Organisation of audio-visual three-dimensional space

Marina Zannoli

ORGANISATION OF AUDIO-VISUAL THREE-DIMENSIONAL SPACE

PhD Thesis, Université Paris Descartes

ORGANISATION DE L'ESPACE AUDIOVISUEL TRIDIMENSIONEL

Thèse de Doctorat, Université Paris Descartes

Marina Zannoli

Directeur: Pascal Mamassian

Membres du Jury:
Rapporteurs: Pr Julie Harris (University of St Andrews), Dr Eli Brenner (Vrije Universiteit)
Examinateurs: Dr Wendy Adams (University of Southampton), Pr Patrick Cavanagh (Université Paris Descartes), Dr Jean-Baptiste Durand (CNRS)

Date de soutenance: 28 septembre 2012

Laboratoire Psychologie de la Perception
ED 261: Cognition, Comportement, Conduites Humaines
Université Paris Descartes, Sorbonne Paris Cité et CNRS UMR 8158

Abstract

Stereopsis refers the perception of depth that arises when a scene is viewed binocularly. The visual system relies on the horizontal disparities between the images from the left and right eyes to compute a map of the different depth values present in the scene. It is usually thought that the stereoscopic system is encapsulated and highly constrained by the wiring of neurons from the primary visual areas (V1/V2) to higher integrative areas in the ventral and dorsal streams (V3, inferior temporal cortex, MT). Throughout four distinct experimental projects, we investigated how the visual system makes use of binocular disparity to compute the depth of objects. In summary, we show that the processing of binocular disparity can be substantially influenced by other types of information such as binocular occlusion or sound. In more details, our experimental results suggest that:

(1) da Vinci stereopsis is solved by a mechanism that integrates classic stereoscopic processes (double fusion), geometrical constraints (monocular objects are necessarily hidden to one eye, therefore they are located behind the plane of the occluder) and prior information (a preference for small disparities).

(2) The processing of motion-in-depth can be influenced by auditory information: a sound that is temporally correlated with a stereomotion-defined target can substantially improve visual search.
Stereomotion detectors are optimally suited to track 3D motion but poorly suited to process 2D motion.

(3) Grouping binocular disparity with an orthogonal auditory signal (pitch) can increase stereoacuity by approximately 30%.

Key words: stereopsis, da Vinci stereopsis, stereomotion, visual search, audio-visual integration, stereoacuity.

Résumé

Le terme stéréopsie renvoie à la sensation de profondeur qui est perçue lorsqu'une scène est vue de manière binoculaire. Le système visuel s'appuie sur les disparités horizontales entre les images projetées sur les yeux gauche et droit pour calculer une carte des différentes profondeurs présentes dans la scène visuelle. Il est communément admis que le système stéréoscopique est encapsulé et fortement contraint par les connexions neuronales qui s'étendent des aires visuelles primaires (V1/V2) aux aires intégratives des voies dorsales et ventrales (V3, cortex temporal inférieur, MT). A travers quatre projets expérimentaux, nous avons étudié comment le système visuel utilise la disparité binoculaire pour calculer la profondeur des objets. Nous avons montré que le traitement de la disparité binoculaire peut être fortement influencé par d'autres sources d'information telles que l'occlusion binoculaire ou le son. Plus précisément, nos résultats expérimentaux suggèrent que :

(1) La stéréo de da Vinci est résolue par un mécanisme qui intègre des processus de stéréo classiques (double fusion), des contraintes géométriques (les objets monoculaires sont nécessairement cachés à un œil, par conséquent ils sont situés derrière le plan de l'objet caché) et des connaissances à priori (une préférence pour les faibles disparités).

(2) Le traitement du mouvement en profondeur peut être influencé par une information auditive : un son temporellement corrélé avec une cible définie par le mouvement stéréo peut améliorer significativement la recherche visuelle.
Les détecteurs de mouvement stéréo sont optimalement adaptés pour détecter le mouvement 3D mais peu adaptés pour traiter le mouvement 2D.

(3) Grouper la disparité binoculaire avec un signal auditif dans une dimension orthogonale (hauteur tonale) peut améliorer l'acuité stéréo d'approximativement 30%.

Mots-clés: stéréopsie, stéréo de da Vinci, mouvement stéréo, recherche visuelle, intégration multisensorielle, acuité stéréo.

Acknowledgements

First, I would like to deeply thank the members of my PhD committee for giving me the honour to evaluate my work.

Second, I am extremely grateful to Pascal Mamassian for giving me the opportunity to work under his supervision during the past five years. He was able to make time for me in his incredibly busy schedule (including emails at one in the morning or during the weekends). I thank him for sharing his bright ideas with me. He suffered my twisted humour during five years with bravery and patience.

Along with Pascal are the members of the LPP lab in Paris. I am thankful to Adrien Chopin, Simon Barthelmé, Thomas Otto, Mark Wexler, Trevor Agus, Victor, Francis and Agnès Léger for their constructive comments on my work.

A very special thank you goes to Marie de Montalembert and Agnès Léger. You have been the most supporting friends in the world.

Then, come the Australians. I think that my visit in David Alais's lab would have been far less rewarding if I haven't met Emily Orchard-Mills, Susan Wardle and Johahn Leung. You have been so welcoming and have run so many of my experiments. Johahn: a special thank you for proofreading! Of course, I am deeply thankful to David Alais and John Cass for their time, intelligence and enthusiasm in our collaborations.

I am also very thankful to Katharina Zeiner, Inna Tsirlin and especially Laurie Wilcox for hours spent discussing why monocular occlusions are so fascinating.

I thank my family for their constant support during the past three years.

Finally, I would like to thank Pierre for his help, support, patience and love during this PhD. You're the best!

Articles

- Zannoli, M., & Mamassian, P. (in preparation). The effect of audio-visual grouping on stereoacuity.

- Zannoli, M., Cass, J., Alais, D. & Mamassian, P. (submitted to *Journal of Vision*). Stereomotion detectors are poorly suited to track 2D motion.

- Zannoli, M., Cass, J., Mamassian, P. & Alais, D. (2012) Synchronized Audio-Visual Transients Drive Efficient Visual Search for Motion-in-Depth. *PLoS ONE, 7* (5): e37190.

- Zannoli, M., & Mamassian, P. (2011). The role of transparency in da Vinci stereopsis. *Vision Research, 51*, 2186–2197.

Table of contents

Part 1

Introduction and literature review

I General introduction

"To my astonishment, I began to see in 3D. Ordinary things looked extraordinary. Sink faucets reached out toward me, hanging light fixtures seemed to float in mid-air, and I could see how the outer branches of trees captured whole volumes of space through which the inner branches penetrated. Borders and edges appeared crisper; objects seemed more solid, vibrant, and real. I was overwhelmed by my first stereo view of a snowfall in which I could see the palpable pockets of space between each snowflake."

Sue Barry

Psychology Today

Susan Barry, professor of neurobiology, was stereoblind from birth due to congenital strabismus until she gained stereovision after several years of optometric training. In her book "Fixing My Gaze", "Stereo Sue" describes her first experiences of stereoscopic vision. In an interview given to Psychology Today (see citation above), she tries to capture the ineffable sensation of stereopsis and how it affects our global visual experience. Stereoscopic vision is involved in various complex visual tasks. In her own words, she describes how stereoscopic 3D shape discrimination is used for object recognition (*"I could see how the outer branches of trees captured whole volumes of space through which the inner branches penetrated."*) and guiding of rapid precise actions such as eye movements or hand reaching (*"Sink faucets reached out toward me."*). She also explains how the acute sensitivity of the stereoscopic system to depth discontinuities allows fine object segmentation (*"borders and edges appeared crisper"*). By referring to the spatial configuration of snowflakes (*"I could see the palpable pockets of space between each snowflake"*), Sue Barry gives a practical example of the extraordinary acuity of the stereoscopic system.

The impact of Sue Barry's book on the scientific community was ultimately substantial but lukewarm at first. Over forty years ago, Hubel & Wiesel (1962) demonstrated the existence of a critical period in the development of the visual system during which equal binocular inputs are necessary of normal

development of cortical and perceptual binocularity. Their discovery was based on induced strabismus in kittens. If caused during the first days of life, it resulted in massive loss of binocular cells in the primary visual cortex. Cortical columns of neurons (Fig. I.1) normally receiving inputs from the two eyes were instead activated only by the healthy eye. Ocular dominance columns connected to the strabismic eye were small and columns connected to the non-deviating eye abnormally large. This unequal ocular dominance distribution was still found after the three-months critical period.

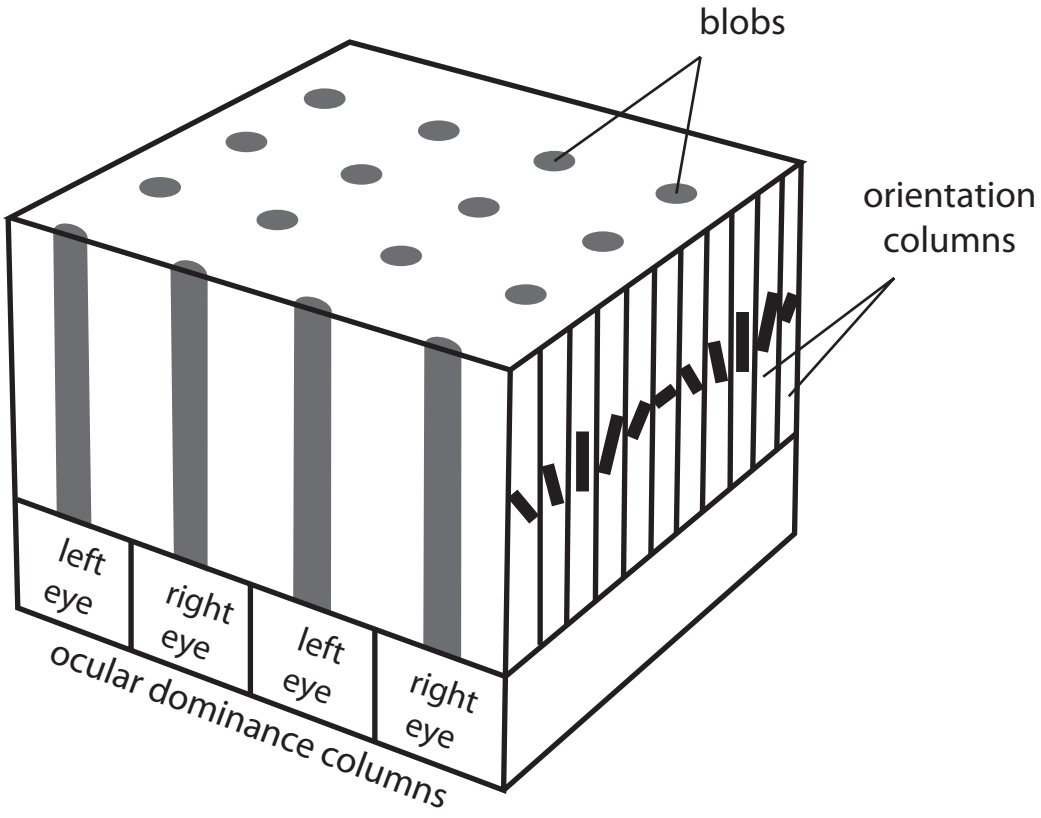


Figure I.1 | Normal ocular dominance columns in the primary visual cortex. Each point in the visual field produces a response in a 2x2 mm area of the primary visual cortex called a hypercolumn. Each of these areas contains two pairs of ocular dominance columns. Within one ocular dominance column, an alternation of blobs and interblobs contains neurons sensitive to all possible orientations across 180°

In 1981, Hubel & Wiesel were awarded a Nobel prize for their work on the development of the visual system and the description of ocular dominance

columns. Since then, it was accepted truth that a critical period of normal binocular input is required for healthy stereoscopic development. As a result, congenital strabismic patients never received optometric rehabilitation.

The publication of Sue Barry's book was closely followed by an article by Ding & Levi (2011) reporting that human adults with abnormal binocular vision (due to strabismus or amblyopia) recovered stereopsis through perceptual learning. Stereopsis, the same visual attribute used over forty years ago to demonstrate the existence of a critical period for the visual system, now bears striking evidence of functional plasticity. Because it is highly dependent on the wiring of neurons spread throughout several regions of the visual cortex and because it is involved in a significant number of various visual tasks, stereopsis can be considered as a canonical representation of visual processing.

Lately, the study of stereopsis has benefited from the recent development of 3D movies, television and 3D gaming consoles that have drawn attention to specific issues such as the vergence-accommodation conflict or visual plasticity.

Throughout the introduction of this thesis, we will first briefly introduce the basic concepts of binocular vision (fusion, binocular summation and binocular rivalry) and then move on to a more detailed review of stereopsis. The purpose of the literature review on stereopsis is to give a broad overview of the current knowledge on the field, highlight apparent contradictions and stress unsolved issues using results from the psychophysics, neurophysiology, imaging and modelling literature. The experimental work conducted during the past three years is detailed in the three experimental chapters. Each chapter comprises an Introduction section followed by an experimental report in the form of a scientific article. The goal of these Introduction sections is to give a critical review of the literature on the topic of the studies presented in each chapter and present the issue addressed in the study. In the second chapter, we present a series of experiments on the role of monocular regions in stereoscopic processing. In the third chapter we present two experimental projects on the processing of motion-in-depth. In the fourth chapter, we describe a series of experiments on auditory facilitation of stereoacuity. Finally, in the General

discussion and Conclusion sections we discuss altogether the results obtained in the four experimental projects presented in this thesis.

II Binocular vision

1 History

By means of mathematics and individual introspection, the ancient Greeks were among the first to expound theories about the optics of the eyes and the transformation of light into visual percepts. Around the 5th century BC, the distance of an object was thought to be sensed by the length of the light rays arriving to the eyes. The first mention of binocular disparity was made by Aristotle (384-322 BC). He realized that one sees double when an object does not fall on corresponding points in the two eyes, for example as a result of misconvergence. Euclid (323-285 BC) was the first to suggest a potential role of occlusion geometry in spatial perception. He observed that a far object is occluded by a nearer object by a different extent in the two eyes and therefore that two eyes see more of an object than either eye alone when the object is smaller than the interocular distance. Ptolemy (c. AD 100-175) hypothesized that binocular vision is used to actively bring the visual axes onto the object of interest, making the first mention of vergence eye movements. Based on anatomical observations, Galen (c. AD 129-201) proposed that the combination of the optic nerves in the chiasma unites impressions from the two eyes.

Almost one century later in Egypt, Alhazen (c. AD 965-1040) confirmed that the movements of the eyes are conjoint to converge on the object of interest. He also explained that the lines of sight for objects close to the intersection of the visual axes fall on corresponding points of the two retinas.

Interest in visual perception was lost during six centuries and regained in Europe by the end of the middle ages. Based on previous observations from the Greeks, artists such as da Vinci (1452-1519) became interested in the issue of

representing three-dimensional space into pictorial space. Da Vinci demonstrated that what can be seen from two vantage points cannot be faithfully represented on a canvas. He also reported that an object occludes a different part of the scene to each eye and that occlusion disparity can be a source of information to depth. Descartes (1596-1650) extended Galen's conclusions and hypothesized that the united image from the two eyes is projected back onto the brain (on the pineal gland, Fig. II.1).

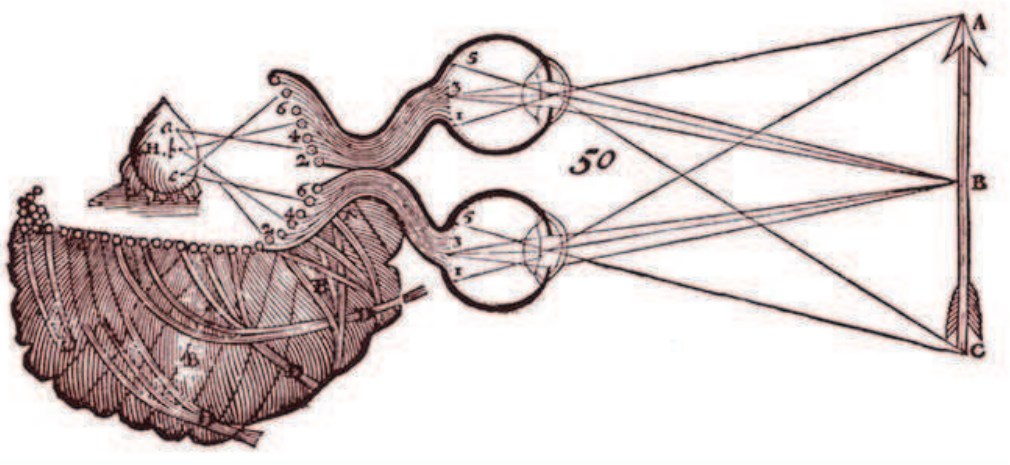


Figure II.1 | Illustration of the stereoscopic visual system by Descartes. Corresponding points of the arrow are projected upon the surface of the cerebral ventricles and then to the pineal gland, H ("seat of imagination and common sense"). (reproduced from Polyak, 1957)

Furthermore, Descartes and Rohault (1618-1672) made the first reference to retinotopy by suggesting that corresponding points in the retina are spatially mapped onto the pineal gland. This assumption was enriched with Newton's (1642-1727) proposition that visual paths are segregated: the temporal half of the retina is treated ipsilaterally while the nasal part is treated contralaterally. Prévost (1751-1839) was the first to describe the horopter (locus of points in space that can be correctly fused and yield single vision) whose geometry was established by Vieth and Müller a few years later.

In 1838, Wheatstone designed the first mirror stereoscope (Fig. II.2) and demonstrated that binocular disparity (horizontal separation between the

projections of an object's image in the left and right eyes) plays a crucial role in depth perception.

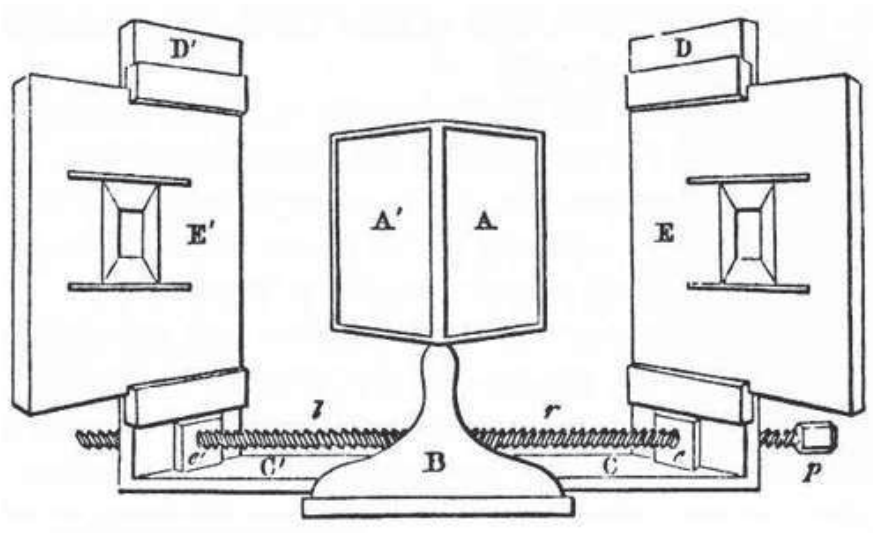


Figure II.2 | Illustration of Wheatstone's first mirror stereoscope. (reproduced from Wheatstone, 1838)

Before 1960, it was believed that stereopsis is the product of high-level cognitive processes. According to Helmholtz (1821-1894) and his student Wundt (1832-1920), a united image of the world was produced by a "mental act" and not by "any anatomical process". The existence of neurons sensitive to binocular inputs was first suggested by Ramon & Cajal in 1911 and then demonstrated by Hubel & Wiesel (1959; 1962). A few years later, Pettigrew, an undergraduate student, recorded cells sensitive exclusively to binocular disparity in the Cat's cortex in the University of Sydney (Pettigrew, Nikara, & Bishop, 1968) and in the University of Berkeley (Barlow, Blakemore, & Pettigrew, 1967). This provided the first evidence of the existence of disparity detectors.

At the same time, Julesz (1964a) used random-dot stereograms (RDSs — pairs of images of random dots which produce a sensation of depth when seen separately by the two eyes) to demonstrate that binocular disparity is sufficient for the perception of depth. RDSs were then used by Marr & Poggio (1979; 1976) to develop the first algorithm capable to solving stereoscopic depth exclusively on the basis of binocular disparity. (For an exhaustive review on the history of binocular vision, see Howard, 2002).

2 Fusion of binocular images

By the time Wheatstone demonstrated the importance of binocular disparity in depth perception, there co-existed two theories of how binocular images are combined into a single percept. In the *fusion* theory, similar images that fall on corresponding points of the retinas access the visual system simultaneously and are fused to form a unitary percept while dissimilar images are suppressed alternatively. According to the *suppression* theory, both similar and dissimilar images engage in alternating suppression at an early stage of visual processing. The discovery of binocular cells in the striate cortex of the cat by Hubel & Wiesel (1962) favoured the idea that the fusion of similar images happen at a low level of processing and fusion became the prevailing theory.

The fusion of binocular images brings several advantages in addition to stereoscopic vision. For example, complex visual tasks such as reading or visuo-motor coordination are better with binocular viewing even if the visual stimuli do not contain any stereoscopic depth information (R. K. Jones & Lee, 1981; Sheedy, Bailey, Buri, & Bass, 1986). As we will see in the following section, detection and discrimination of visual stimuli are better when performed by two eyes instead of one. This phenomenon is called *binocular summation*. However, when images are too different they compete for access to higher levels of visual processing, resulting in alternating perception of the two. This phenomenon is called *binocular rivalry*. In the last section, we will overview the main issues concerning binocular rivalry: what rivals during rivalry, what triggers alternation and what survives suppression. The mechanisms underlying stereoscopic vision will be the subject of a separate chapter of this introduction.

3 Binocular summation

Binocular summation refers to the process by which binocular vision is enhanced compared to what would be expected with monocular viewing. Binocular summation results in increased sensitivity in detection and discrimination tasks. For example, Blake & Fox (1973) showed that visual

resolution measured with high-contrast gratings was slightly higher with binocular vision.

Different causes for binocular summation have been suggested. First, a series of psychophysical studies reveal that low-level factors can contribute to binocular summation. For example, it has been shown that pupil size in one eye is influenced by illumination in the other eye, suggesting that subcortical centres that control pupillary dilatation combine inputs from the two eyes (Thomson, 1947). Increased binocular acuity could also be due to binocular fixation being steadier.

Apart from low-level facilitation, binocular summation is thought to be the main product of probability summation. There is a statistical advantage of having two detectors (eyes) instead of one. Between the sixties and the eighties, there were two alterative accounts of probability summation, both assumed that binocular summation was achieved through a single channel and posited a summation ratio of 40% between monocular and binocular thresholds. Campbell & Green (1965) proposed that monocular signals are linearly summed and that the signal-to-noise ratio is decreased because the two sources of noise are uncorrelated. Alternatively, Legge (1984a; 1984b) posited that the binocular contrast of a grating is the quadratic sum of the monocular contrasts. Monocular signals are squared prior to combination. Anderson & Movshon (1989) used adaptation and noise to refute the single-channel assumption and proposed that there are several ocular-dominance channels of binocular summation. The maximum summation ratio of 40% was then questioned by several studies that found substantially larger summation ratios (Meese, Georgeson, & Baker, 2006).

More recent multi-stage models of binocular summation have been proposed. For example, the models by Ding and Sperling (2006) and Meese, Georgeson, & Baker (2006) are based on contrast gain control mechanisms before and after combination of the two monocular signals.

4 Binocular rivalry

When the images arriving to the two eyes are too dissimilar in colour, orientation, motion, etc., the visual system fails to fuse them into a single coherent percept. The images from the two eyes then rival for dominance and access to perceptual awareness, and the observer's perception alternates every few seconds between one image and the other (Fig. II.3).

Various aspects of the visual stimulation are known to influence binocular rivalry. For example, Levelt (1965; 1966) proposed that the strength of a stimulus determines the duration of its suppression: the weaker it is the longer it is suppressed. He proposed that the strength of a stimulus is proportional to the density of contour in the image. Mueller & Blake (1989) later showed that the contrast of rival patterns had an effect on the rate of alternation. Blur is also known to affect binocular rivalry: Humphriss (1982) demonstrated that defocussed images tend to be suppressed in favour of sharp images.

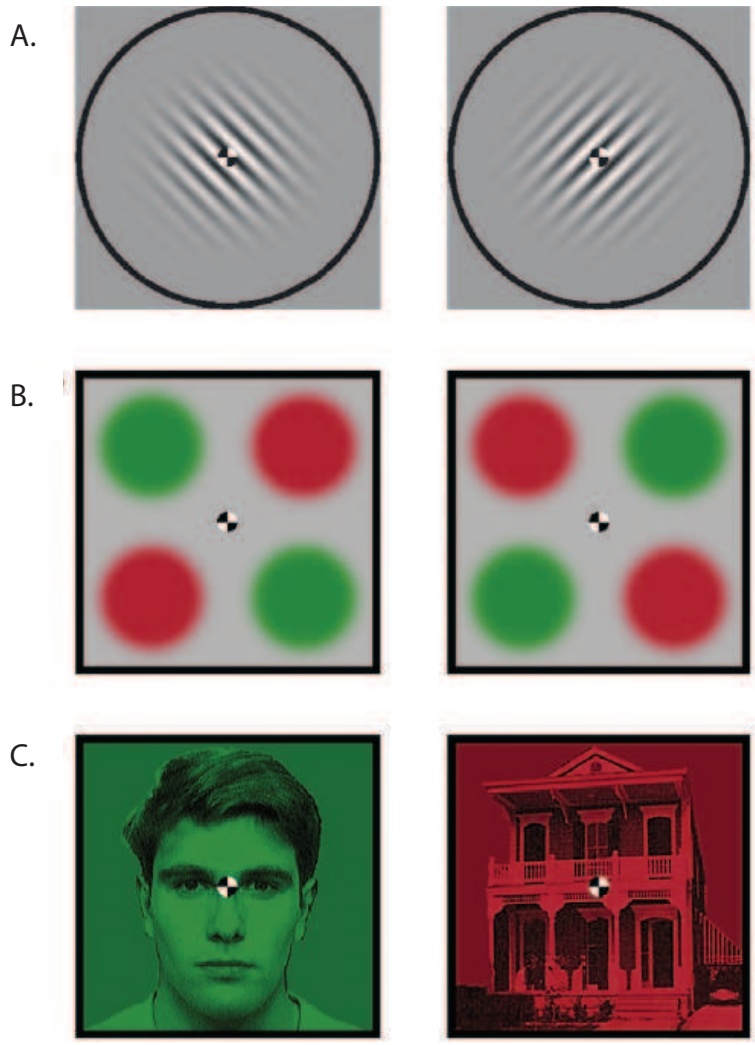


Figure II.3| Examples of binocular rivalry stimuli. The left and right columns show images presented to the left and right eyes respectively. A. Dichoptic orthogonal gratings. B. Stimuli used to study interocular grouping, adapted from Tong, Nakayama, Vaughan, & Kanwisher (1998). C. Rivalry using complex objects, adapted from Kovács, Papathomas, Yang, & Fehér (1996). (reproduced from Tong, Meng, & Blake ,2006)

4.1 Eye- versus pattern-rivalry

Traditionally, two alternative conceptions of binocular rivalry co-existed until the mid-nineties. According to one view, competition occurs between neurons in the primary visual cortex (Blake, 1989; Tong, 2001) or in the lateral geniculate nucleus (Lehky, 1988) that represent local corresponding regions in the two eyes. Alternatively, binocular rivalry could take place in later stages of

visual processing and reflect competition between incompatible patterns (e.g. Diaz-Caneja, 1928; Kovács et al., 1996) that could be distributed between the two eyes (Fig. II.4).

Figure II.4 | Eye- versus pattern-rivalry. When composite images as seen in the lower pair of images are presented to the left and right eyes, perception alternates between the two coherent percepts shown in the upper pair of images. (reproduced from Kovács, Papathomas, Yang & Fehér, 1996)

More recently, models incorporating elements of both views have been proposed, promoting the idea that rivalry is based on neural competition at multiple stages of visual processing (Freeman, 2005; Wilson, 2003). Neural competition is mediated by reciprocal inhibition between visual neurons. A group of neurons dominates temporarily until they can no longer inhibit the activity of competing neurons. When inhibition breaks down, perceptual dominance is reversed. This competition is thought to take place both between monocular and pattern-selective neurons (Fig. II.5).

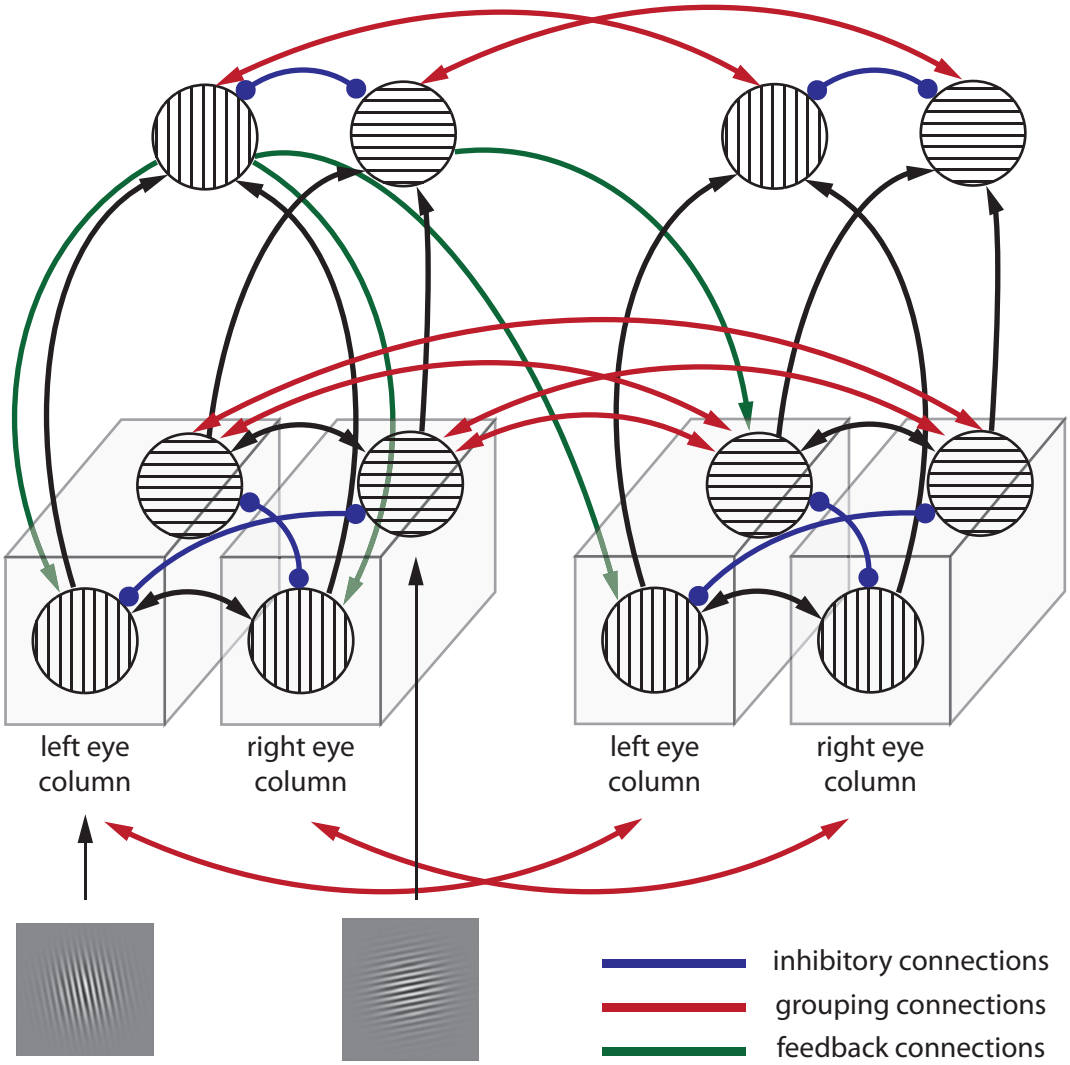


Figure II.5 | Schematic diagram of inhibitory and excitatory connections in a hybrid rivalry model. Reciprocal inhibitory connections between monocular neurons and binocular neurons (blue lines) account for eye-based and pattern-based visual suppression, respectively. Reciprocal excitatory connections (red lines). These lateral interactions might account for eye-based grouping, low-level grouping between monocular neurons with similar pattern preferences including interocular grouping, and high-level pattern-based grouping between binocular neurons. Excitatory feedback projections (green lines) might account for top-down influences of visual attention and also feedback effects of perceptual grouping. (adapted from Tong et al., 2006)

4.2 Perceptual transitions in binocular rivalry

There is a consensus around the idea that alternations in binocular rivalry are mainly the product of adaptation. The activity of neurons associated with the dominant percept progressively vanishes over time, reducing the strength of its inhibition on the suppressed group of neurons. This dynamic process eventually leads to a reversal in the balance of activity between the two neural representations (Alais, Cass, O'Shea, & Blake, 2010; Blake, Sobel, & Gilroy, 2003). Since adaptation takes place at all stages of visual processing, this hypothesis is compatible with both eye- and pattern-rivalry.

However, adaptation cannot fully account for the dynamics of binocular rivalry. Incorporating neural noise either in the inhibitory or the excitatory network has been proposed to explain the stochastic properties of rivalry alternations (van Ee, 2009). Attention has been found to bias the first percept and the duration of subsequent alternation sequences (Chong, Tadin, & Blake, 2005). Recently, Chopin & Mamassian (2012) demonstrated that the current percept in binocular rivalry is strongly influenced by a time window of stimuli presented remotely in the past. They proposed that the remote past is used to estimate statistics about the world and that the current percept is the one that matches these statistics.

4.3 Effects of suppressed images

fMRI recordings have shown that activation evoked by the suppressed stimulus is reduced compared to the activation produced by the dominant image. However, various psychophysical paradigms have demonstrated that suppressed stimuli can affect visual processing. For example, it has been shown that suppressed stimuli can induce adaptation aftereffects, visual priming (Almeida, Mahon, Nakayama, & Caramazza, 2008) and covertly guide attention to definite locations of the suppressed image (Jiang & He, 2006). It has also been shown that stimuli that convey meaningful or emotional information are suppressed for a shorter duration (Jiang, Costello, & He, 2007).

4.4 Binocular rivalry in the brain

Imaging techniques such as EEG or fMRI have been used to investigate the neural correlates of the inhibitory components and reversals in binocular rivalry. fMRI techniques have allowed researchers to tag the activity corresponding the each of the two percepts involved in the alternation. For example, Tong and colleagues (1998) induced rivalry between face and house pictures and showed that activation in the regions selectively sensitive to these two categories was correlated with the dynamics of rivalry.

As explained in the first pages of this section, binocular rivalry can be seen as a failure in fusing the images from the two eyes. A majority of the computational models of stereoscopic processing has focused on the computations taking place once fusion is achieved. A few alternative models have intended to include binocular rivalry as part of the resolution of the correspondence problem. One exception is Hayashi, Maeda, Shimojo, & Tachi (2004) who proposed that rivalry is the default outcome of the system when binocular matching fails (see chapter IV, section 1.5 for a more detailed review of this type of stereo models).

5 Binocular rivalry and stereopsis

According to the parallel pathways theory (Wolfe 1986, Kaufman 1964), stereopsis and binocular rivalry are processed in separate pathways. In particular, Wolfe argued that suppression is active in the rivalry pathway at all times, even when the two monocular views are identical. In parallel, the suppressed image is used to compute binocular disparity. In favour of this theory, Kaufman (1964) showed that a random-dot stereogram containing binocular disparities is seen in depth while the background (with a different colour in the two eyes' images) is seen as rivalrous (Fig. II.6). Following this framework, Carlson & He (2000) proposed that the chromatic parvo-cellular pathway deals with binocular rivalry while the achromatic magno-cellular pathway extracts binocular disparity. However, there is currently no convincing

evidence that these two pathways (hence processes) are genuinely parallel and not sequential. It remains to be demonstrated that stereoscopic vision and binocular rivalry can be based on the same substrate.

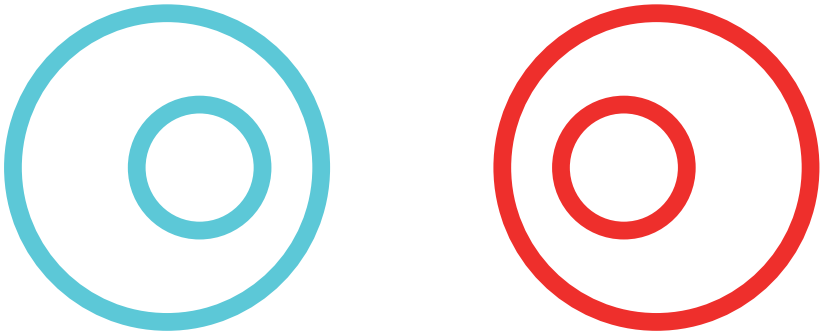


Figure II.6 | Colour rivalry in stereoscopic vision. Fusing these two images creates relative depth between the two embedded circles and colour rivalry at the same time. (adapted from Treisman, 1962)

Today, the predominant theory (Blake, 1989; Julesz & Tyler, 1976) advances that fusion is the first step and that the extraction of binocular disparity takes place only if fusion is successful. When fusion fails, images a locally engaged in the second step, which is binocular rivalry. It is worth noting that unpaired regions of an image (seen by one eye only) do not engage in rivalry or suppression when they are consistent with the geometry of occlusion present in the scene (Nakayama & Shimojo, 1990). See chapter IV for a detailed review and an experimental study on depth from monocular occlusion.

III Stereopsis

The word stereopsis refers to the impression of depth that arises when a scene is viewed binocularly. The horizontal separation between the eyes creates two different vantage points. The images seen by the two eyes are therefore slightly different. These differences are called binocular disparities (Fig. III.1) and they are used by the visual system to recover the depth position of the objects and surfaces present in the visual scene as well as their 3D structure.

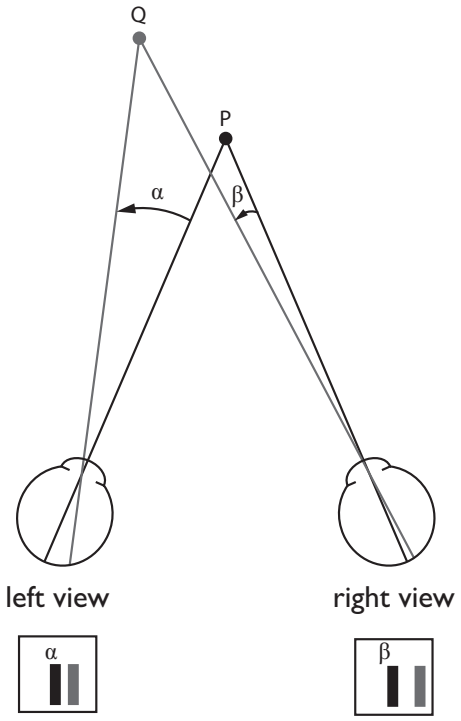


Figure III.1 | Top down view of the two eyes fixating point P. The relative depth between points P and Q is computed from the angular disparity = α - β.

In the present section, we will give a brief overview of the knowledge acquired on stereopsis since the nineteenth century. First, we will focus on the basic properties of the stereoscopic system, referring mainly to psychophysical studies. Then we will rely on neurophysiological and imaging studies to try to understand how binocular disparity is processed in the brain. Finally, we will outline the main computational

and biologically-inspired concepts used to model the processing of binocular disparity in the stereoscopic system

1 Stages of stereoscopic processing

In order to precisely evaluate the depth of objects and surfaces, the visual system relies on outputs from neurons sensitive to such basic properties as orientation and spatial frequency. As we will see, the visual system will be confronted by several computational problems to transform these outputs into complex depth maps. Backus, Fleet, Parker & Heeger (2001) identified six stages of stereoscopic processing. The first three stages are involved in the computation of disparity maps based on retinal disparity inputs. Once absolute disparities (relative to the point of fixation) are detected, they are converted into relative disparities (independent of fixation). Several psychophysical studies have shown the importance of relative disparity for stereopsis. For example, it has been shown that changes in absolute disparity do not produce changes in perceived depth (Erkelens & Collewijn, 1985) and that stereoscopic thresholds are not a simple function of absolute disparity (Andrews, Glennerster, & Parker, 2001). Disparity information is then spread across the surface to *fill-in* ambiguous areas and construct the disparity map. This process is also known as *disparity interpolation* (Warren, Maloney, & Landy, 2002; 2004). The fourth stage is segmentation based on disparity (Westheimer, 1986) were the disparity map is segmented into discrete objects. The fifth stage is the disparity calibration in order to estimate depth, where disparity values are scaled by viewing distance to extrapolate the actual depth between different surfaces. Finally, the percept created by stereopsis can drive attention to specific locations of space (He & Nakayama, 1995).

2 Spatial and temporal limits of stereopsis

To construct a representative map of the disparities present in a scene, the stereoscopic system must solve the "correspondence problem". It has to detect the corresponding points in the two eyes' images and discard potential false matches. The

possible solutions to the correspondence problem are constrained by various spatial and temporal limits of the stereoscopic system.

2.1 Spatial limits of stereopsis

2.1.1 The horopter, the Vieth-Muller circle and Panum's fusional area

Aguilonius introduced the term *horopter* in 1613 to describe the location in space in which fused images appear to lie. Two hundred years later, Vieth and Müller argued from geometry that the theoretical horopter should be a circle (now known as the *Vieth-Müller circle*) passing through the point of fixation and the centres of the eyes. When measured empirically, the horopter is found to be flattened compared to the Vieth-Müller circle. The detection of planarity constitutes a challenge for the stereoscopic system and it has been suggested that there exists a prior for perceiving fronto-parallel planes rather than curved surfaces.

If defined by singleness of vision (fusion), the empirical horopter is much thicker. This range of disparities within which fusion is achieved has been studied by Panum (1858) and called the *Panum's fusional area* (Fig. III.2). The Panum's fusional area expands around the empirical horopter. Stimuli containing disparities outside this range lead to diplopic images. Ogle (1952) measured the maximum disparity (d_{max}) that produced depth with fused images (± 5 arcmin), depth with double images (± 10 arcmin) and vague impression of depth with diplopia (± 15 arcmin). He dubbed the first two *patent stereopsis* and the last *qualitative stereopsis*. It is worth mentioning that more recent studies have found larger estimates of these critical values.

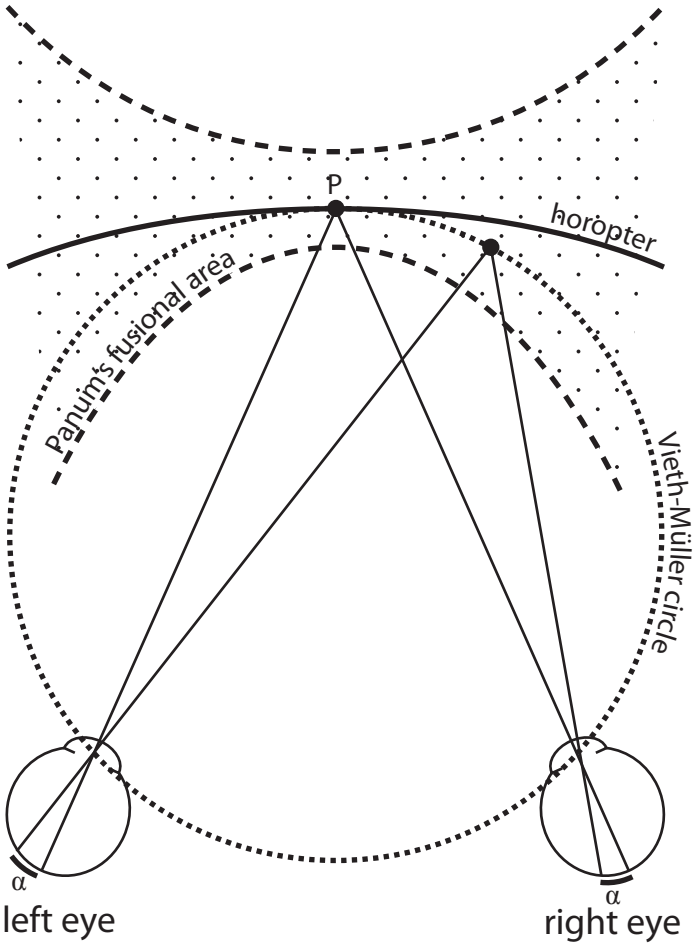


Figure III.2 | Schematic representation of the geometry of stereopsis. Top down view of the two eyes fixating point P. The horopter, the Vieth-Müller circle and the Panum's fusional area. Two points falling on the Vieth-Müller circle project on corresponding points of the two retinas and therefore subtend the same angle (α).

2.1.2 Stereoacuity

Stereoacuity is the smallest detectable depth difference between two stimuli when binocular disparity is the only cue to depth. The first stereoacuity test was developed by Helmholtz: a vertical rod had to be adjusted in depth to appear in the same plane as two flanking rods. Later, the Howard-Dolman test in which observers had to judge the depth of one rod relative to another was used by the American Air Force on pilots and demonstrated that stereoacuity can be as fine as 2 arcsec (see chapter VI for an experimental application of this method). In 1960, Julesz used random-dot stereograms (RDSs, Fig. III.4 & III.5) to measure stereoacuity in the absence of any monocular depth cue (such as perspective, blur or motion parallax). To create a RDS,

pixels of an array are randomly selected to be black or white. When the same RDS image is presented to the two eyes, a flat plane is perceived. If a portion of one of the two images is copied onto the other with a lateral displacement, it is perceived as a surface floating in depth. The distance between this surface and the plane of the image is determined by the amount of lateral displacement. Julesz found that stereoacuity from RDSs was highly accurate even though they took longer to see. RDSs were later used in standardized Stereoacuity tests such as the TNO test.

Figure III.3 | Stereo pair which, when viewed stereoscopically, contains a central rectangle perceived behind. (Reproduced from Julesz, 1964).

1	0	1	0	1	0	0	1	0
1	0	X	A	A	B	B	0	0
0	0	Y	B	A	B	A	1	1
0	1	0	0	1	1	1	0	1
1	1	0	0	0	1	1	0	0
0	0	1	0	1	0	0	1	0
1	1	0	1	0	1	1	0	0
1	0	A	A	B	A	X	0	1
1	1	B	B	A	B	X	1	0
0	1	0	0	0	1	1	1	1

1	0	1	0	1	0	0	1	0
1	0	A	A	B	B	Y	0	0
0	0	B	A	B	A	X	1	1
0	1	0	0	1	1	1	0	1
1	1	0	0	0	1	1	0	0
0	0	1	0	1	0	0	1	0
1	1	0	1	0	1	1	0	0
1	0	Y	A	A	B	A	0	1
1	1	Y	B	B	A	B	1	0
0	1	0	0	0	1	1	1	1

Figure III.4 | Illustration of the method by which the stereo pair of Fig. 4 was generated. Rectangle sectors of the left image were shifted either to the left of the right to create disparity between the two images. Positive disparity was added to the lower rectangle, negative disparity was added to the upper one. (Reproduced from Julesz, 1964).

Stereoacuity has been found to be highly dependent on several aspects of the stimuli used in the measuring process. For example, when the two test stimuli are presented with a disparity pedestal (with a mean disparity that is different from zero), stereoacuity decreases exponentially with the size of the disparity pedestal (Ogle, 1953).

2.1.3 Stereoresolution

It has also been shown that stereoacuity is scaled by the spatial frequency of the depth modulation in the image. Tyler (1973; 1975) measured spatial *stereoresolution* (the smallest detectable spatial variation in disparity) as a function of spatial frequency by presenting spatially periodic variations in disparity. He found that it was much poorer than the luminance resolution. While the highest detectable spatial frequency for luminance-defined corrugations was about 50 cpd (cycles per degree), it was only about 3 cpd for disparity-defined corrugations. Recent neurophysiological (Nienborg, Bridge, Parker, & Cumming, 2004) and psychophysical (Banks, Gepshtein, & Landy, 2004) results suggest that spatial stereoresolution is limited by the size of the receptive fields of V1 neurons and the type of computations underlying the extraction of disparity (see section 4.4 of this chapter for more details).

2.1.4 Disparity-gradient limit

Burt & Julesz (1980) were the first to mention that the maximum disparity for fusion could be modified by adding nearby objects to the scene. Rather than the Panum's fusional area, these authors proposed that this limit is a ratio, a unitless perceptual constant. This ratio, the *disparity-gradient* (D) between two points is defined by the difference in their disparities (η) divided by the difference between the mean direction (across the two eyes) of the images of one object and the mean direction of the images of the other object (δ) (Fig. III.5). A disparity gradient of zero corresponds to a surface lying on the horopter. When two points are aligned along a visual line in one eye, they have a horizontal disparity gradient of 2 (see Panum's limiting case in chapter IV, section 1.3). This corresponds to the maximum theoretical gradient for opaque surfaces (Trivedi & Lloyd, 1985).

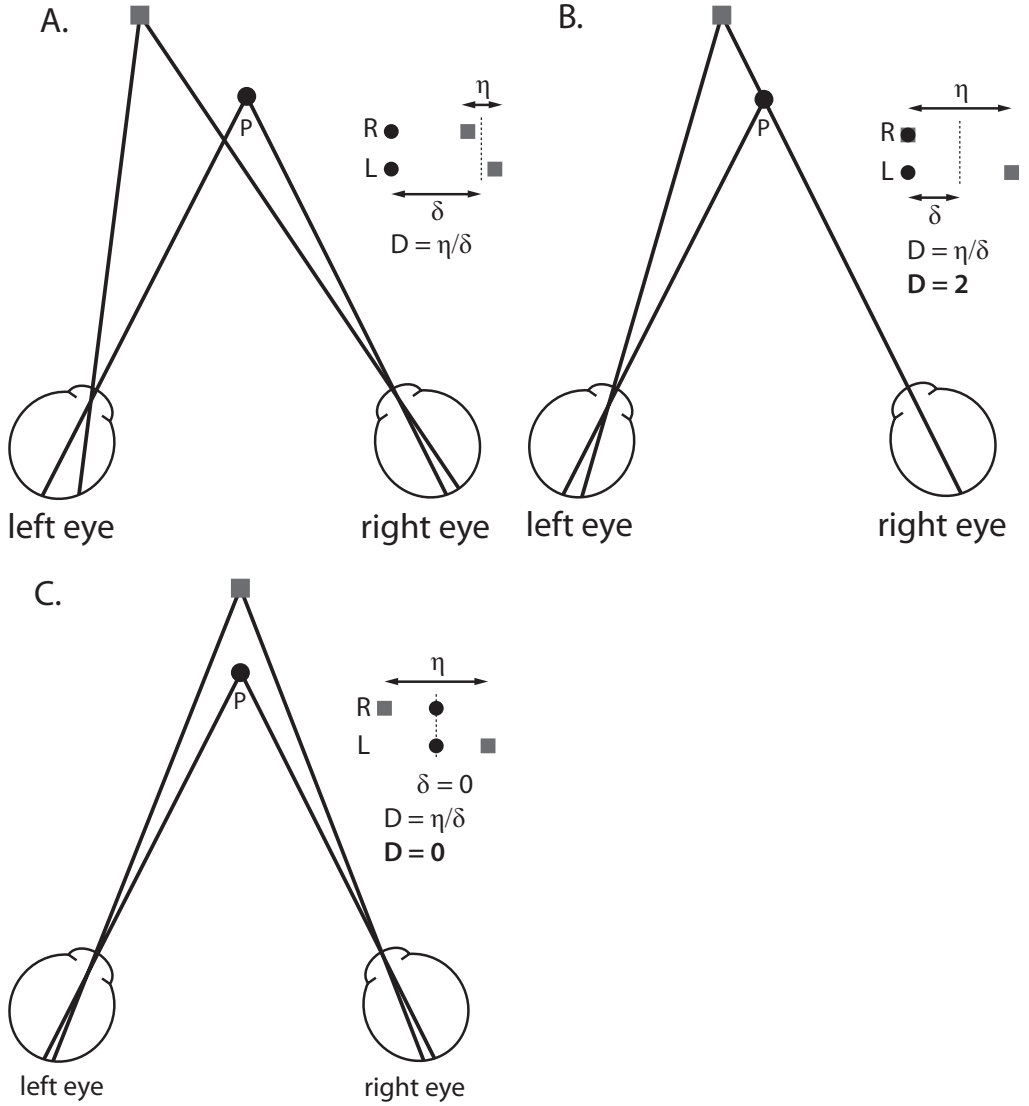


Figure III.5 | Disparity gradients between the black dot and the grey square. The angle η is the difference in disparity between the two objects, δ is the separation in visual angle between the two objects and D is the disparity gradient. A. The two objects have a disparity gradient inferior to 2. B. Illustration of the Panum's limiting case: the two objects are on the same line of sight for one eye. The disparity gradient is 2. C. There is no horizontal separation between the two objects: the disparity gradient is infinite. (redrawn from Howard & Rogers, 2002)

To measure the disparity-gradient limit, Burt & Julesz (1980) systematically varied the vertical separation of two dots and kept the relative disparity between the two constant. They showed that fusion was lost when the disparity-gradient exceeded a critical value of 1.

This critical value of 1 was later incorporated by Pollard, Mayhew & Frisby (1985) in their PMF algorithm for solving the correspondence problem. Recently, Filippini & Banks (2009) proposed that the disparity-gradient limit is a byproduct of estimating disparity by computing the correlations between the two eyes' images (see section 4.4 of this chapter for more details).

2.1.5 Vertical disparity

Vertical disparities are the differences in up-down positions of corresponding points in the left and right eyes images. The size of vertical disparities depends on the orientation of the eyes and the location of the object. The induced effect (Ogle, 1938) constitutes the first clear psychophysical evidence that vertical disparities can convey depth information. He showed that applying a vertical magnification to one eye's image causes the illusion that a frontoparallel surface is rotated about a vertical axis. Objects projected on the eye having the smaller image appear nearer than the objects that are artificially magnified.

Physiological studies on Monkeys have shown that disparity detectors in MT (Maunsell & Van Essen, 1983) and V1/V2 (Durand, Celebrini, & Trotter, 2007; Durand, Zhu, Celebrini, & Trotter, 2002; Gonzalez, Justo, Bermudez, & Perez, 2003) were sensitive to both horizontal and vertical disparities. A more exhaustive review of the physiology of stereopsis can be found in section 3 of this chapter.

Vertical disparity is usually represented by the *vertical size ratio (or VSR)*, which is the ratio of the vertical angles subtended by two points in the left and right eyes. The VSR provides information about the eccentricity of these two points. It increases with eccentricity because the points become closer to one eye and farther from the other. The VSR is also dependent on the absolute viewing distance. As can be seen in Figure III.6, the same VSR can correspond to near points at a small eccentricity or to farther points at a larger eccentricity. VSR therefore provides information about eccentricity at a given distance. If one of the two types of information is known, the other can be deducted.

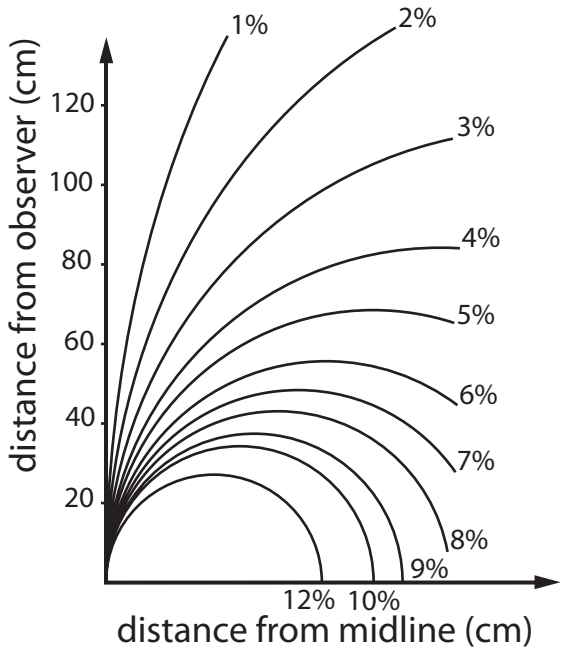


Figure III.6 | Vertical size ratio (VSR) as a function of eccentricity and distance. Each curve connects points of a given scene with the same VSR. The VSR can be the same for an object close to the observer and the medial plane as for an object seen from far away at a large eccentricity. (adapted from Gillam & Lawergren, 1983)

Two theories have been proposed to explain how vertical disparities participate in the solving of the correspondence problem. Mayhew and Longuet-Higgins (1982) postulated that vertical disparities can be used to recover the convergence distance and the angle of eccentric gaze. Alternatively, Gillam & Lawergren (1983) noted that the gradient of VSR as a function of eccentricity is constant for a given viewing distance. Therefore, this VSR gradient can be used to rescale relative disparities when viewing distance cannot be recovered.

More recent psychophysical studies have shown that vertical disparities are used by the visual system to perform various tasks. For example, vertical disparities can be combined with other depth cues for stereoscopic slant perception (Backus & Banks, 1999; Backus, Banks, van Ee, & Crowell, 1999) and vertical disparity discontinuities might be used to detect object boundaries (Serrano-Pedraza, 2010).

2.2 Temporal limits

2.2.1 Stimulus duration

The time of presentation required for perceiving depth from stereopsis greatly varies as a function of the type of stimuli and the experimental procedure used to measure it. Ogle & Weil (1958) were the first to properly measure stereoacuity as a function of stimulus duration with controlled fixation and showed that stereoacuity fell from 10 to 50 arcsec when stimulus duration was reduced from 1 sec to 7.5 ms. It was hypothesized that the integration of disparity over time may be analogous to the integration of luminance. Ogle & Weil's stimuli were luminance-defined rods. Uttal, David & Welke (1994) reported that observers were above chance when asked to recognize a 3D shape on a RDSs presented for 1 ms. The also showed that this performance increased with the number of trials. This effect of practice on the latency of stereopsis for RDSs was also reported by Julesz (1960).

2.2.2 Processing time

In a following study, Julesz (1964a) measured processing time by recording the effect of an unambiguous stereogram on the perception of a following ambiguous one. He found that the inter stimulus interval had to be longer than 50 ms for the first stereogram to bias the perception of the second one. This 50 ms critical value was confirmed by Uttal, Fitzgerald & Eskin (1975) using a masking technique.

2.2.3 Temporal modulation of disparity

Another way of investigating the processing time for stereopsis is to look at the effect of temporal modulations of disparity on stereoacuity. Tyler (1971) compared motion sensitivity for smooth lateral motion and motion-in-depth for sine-wave modulation frequencies from 0.1 Hz to 5 Hz. He showed that sensitivity was best at a modulation frequency of about 1 Hz and that it was substantially better for lateral motion compared to motion-in-depth. Tyler & Norcia (1984) recorded motion perception for RDSs alternating in depth in abrupt jumps and showed that the limit

for apparent depth motion perception was approximately 6 Hz. Above this value, two pulsating planes were perceived simultaneously. A more exhaustive review on motion-in-depth can be found in chapter V, section 1.

3 The physiology of stereopsis

Closely following the discovery of Pettigrew and colleagues (see chapter II), Hubel & Wiesel found similar disparity-selective cells in the area V2 of the monkey's visual cortex. Similar cells were later recorded in the area V1. Poggio and colleagues (1985) found that complex cells in areas V1 and V2 of the monkey respond to binocular disparity embedded in RDSs, providing the first evidence of the existence of cells sensitive exclusively to binocular disparity.

3.1 Disparity detectors

These disparity-selective neurons are now referred to as *disparity detectors*. Each disparity detector is defined by its *disparity tuning function,* which refers to the frequency of firing as a function binocular disparity. The peak of this distribution is the *preferred disparity* and its width indicates the *disparity selectivity* of the neuron. Originally, binocular cells were separated into six categories (Fig. III.7): excitatory cells tuned to zero disparity, tuned inhibitory cells, tuned excitatory cells for crossed disparities, tuned excitatory cells for uncrossed disparities, near cells and far cells (Cumming & DeAngelis, 2001).

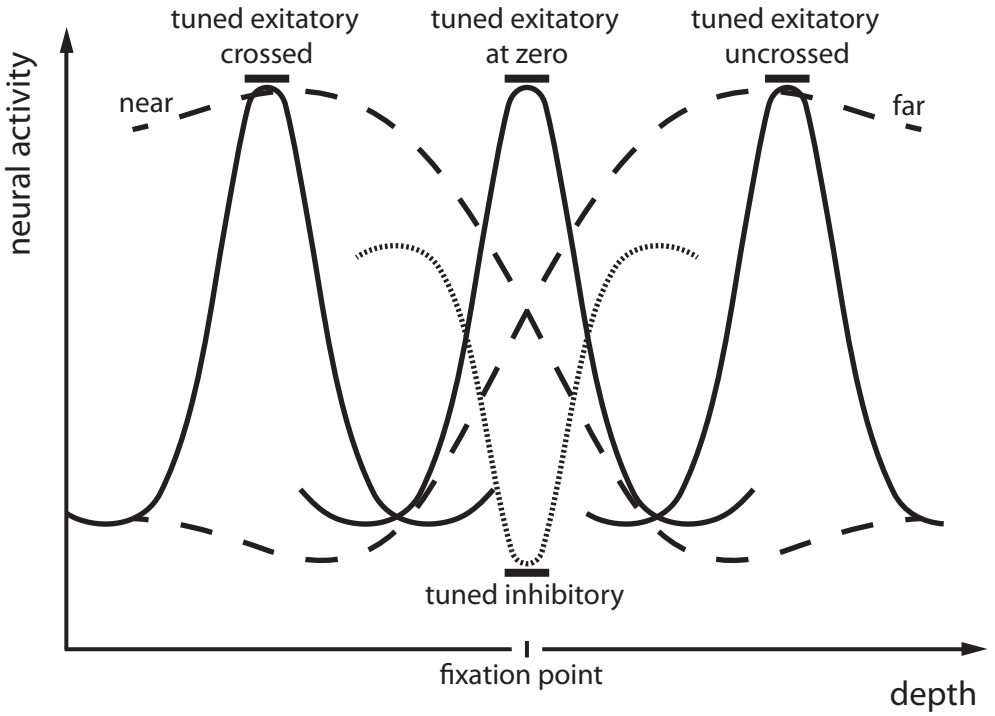


Figure III.7 | Six types of tuning function of disparity detectors. Three types of symmetrical tuned excitatory cells: at zero, crossed, uncrossed. One type of symmetrical tuned inhibitory cell. Two types of asymmetrical near or far cells with broad selectivity. (adapted from Poggio et al., 1985).

This clustering into distinct tuning types was later challenged by other electrophysiological recordings showing a continuous distribution of disparity selectivity (Prince, Cumming & Parker, 2002).

Even though a majority of neurons in the area V1 of the monkey have a preferred disparity, disparity information then undergoes complex transformations in higher visual areas.

3.2 From V1 to V2

There is a body of evidence suggesting that disparity information undergoes a first step of transformations when travelling from V1 to V2. For example, it is hypothesized that V2 is specialized in detecting depth steps and disparity-defined edges (Bredfeldt & Cumming, 2006). While the activity of V1's binocular cells in the monkey appears to be driven exclusively by absolute disparity (Cumming, 1999), some cells in V2 are selective for relative disparity across a range of absolute disparities.

Another study has reported significant choice probabilities in V2 but not V1 in a depth discrimination task (Nienborg & Cumming, 2006). These three examples strongly support the idea that V2 plays a central role in the transformation of binocular disparity into depth information.

3.3 Disparity in the ventral and dorsal streams

Psychophysics, physiology and imaging have now come the consensus that, beyond V2, the processing of disparity is segregated into two main streams that are thought to carry out different types of stereo computation (Fig. III.8): the *ventral stream* (areas from V4 through the inferior temporal cortex) and the *dorsal stream* (areas MT/V5 and MST) (Parker, 2007). This distinction would reflect the specialization of each stream for more general tasks. The ventral stream would be involved in object identification while the dorsal stream would underlie orientation in space and navigation (Goodale & Milner, 1992).

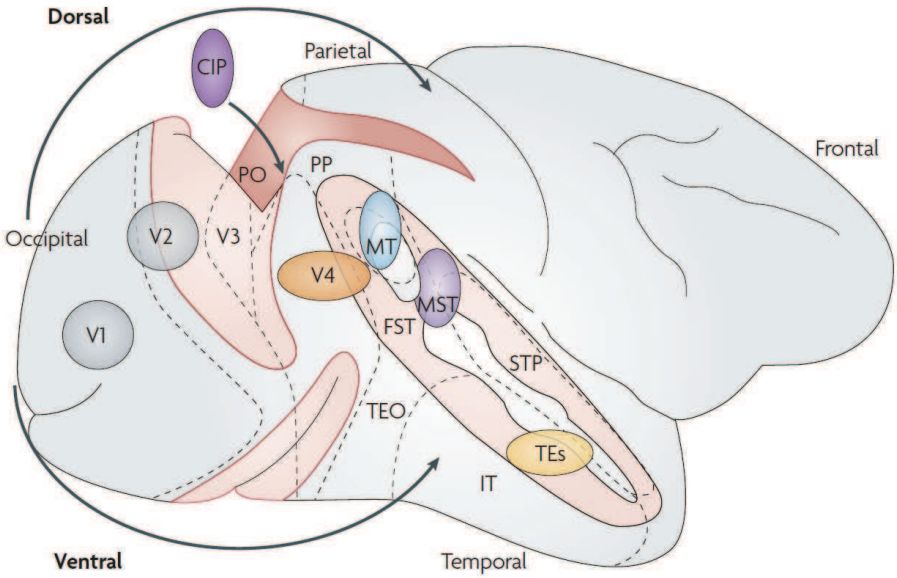


Figure III.8 | Stereovision in the dorsal and ventral pathways. The figure shows a diagrammatic picture of the macaque monkey cortical areas, in which the main flow of visual information through the dorsal and ventral visual pathways is identified by arrows. The ventral visual areas are highlighted with horizontal ellipses of red/orange tints, and the dorsal visual areas are highlighted with vertical ellipses of blue/purple

tints. The early visual areas V1 and V2 are highlighted with neutral grey circles. CIP, caudal intraparietal area; FST, fundal superior temporal area; IT, inferior temporal cortex; MST, medial superior temporal area; MT, medial temporal area; PO, parietooccipital area; PP, posterior parietal cortex; STP, superior temporal polysensory area; TEs, a collection of areas in the anterior inferior temporal cortex. (adapted from Parker, 2007).

Using adaptation and fMRI on humans, Neri, Bridge & Heeger (2004) provided the first evidence of a two-stream dichotomy in humans. They showed that disparity processing relied more on absolute disparity in the dorsal stream while both types of disparity information were preserved in the ventral stream. Inconsistent with Neri and colleagues' findings, Preston, Li, Kourtzi & Welchman (2008) showed that dorsal areas encode disparity magnitude while ventral areas encode disparity sign. Alternatively, these authors suggest that disparity in the ventral stream (area LO) might be used to encode depth configurations and support invariant recognition of objects across different positions in depth. In the dorsal stream, disparity magnitude in areas V3A and V7 might support fine control of body movements while pattern based tuning in hMT+ might be consistent with coarse depth discriminations. Even though the results from Neri et al. and Preston et al. are consistent with a dual pathway dichotomy, they remain conflicting.

3.3.1 The ventral stream

Janssen, Vogels & Orban (2000) provided the first electrophysiological evidence of a specialization for the extraction of 3D shape from disparity in a subregion of the inferior temporal cortex. This finding was backed up by studies showing that the inferior temporal cortex is specifically sensitive to fine depth variations (Uka, Tanabe, Watanabe, & Fujita, 2005). Janssen and colleagues also demonstrated that sensitivity to anticorrelated stereograms (see chapter V, section 1.1.2.1, Fig. V.2) (Cumming & Parker, 1997), found in V1 and MT/V5 & MST was completely abolished in a subregion of the inferior temporal cortex called TE, implying that the correspondence problem is fully solved in the ventral stream (Janssen, Vogels, Liu, & Orban, 2003).

3.3.2 The dorsal stream

The dorsal stream is sensitive to anticorrelated stereograms, suggesting a less elaborated computation of binocular correlation (Janssen et al., 2003). However, electrophysiological recordings in the area MST of monkeys demonstrated that this region plays a central role in driving vergence eye movements. The MT complex has been shown to process motion and disparity (Maunsell & Van Essen, 1983) and more specifically to extract motion-in-depth from changes of disparity over time (Rokers, Cormack, & Huk, 2009) (see chapter V, section 1 for a detailed review on motion-in-depth).

3.3.3 Bridges between the ventral and the dorsal streams

To complement Janssen and colleagues' (Janssen et al., 2000) electrophysiological recordings on the monkey, Chandrasekaran, Canon, Dahmen, Kourtzi & Welchman (2007) measured the correlation between cortical activity (recorded by fMRI) and psychophysical shape judgments. They found that this task was associated with both ventral and dorsal areas, suggesting that the two streams interact to build percepts of 3D shape.

4 Modelling

The challenge for computational models of stereoscopic vision is to be able to determine which parts of an image correspond to which parts of another image. This complex issue is called the correspondence problem (Fig. III.9). Solving the correspondence problem is theoretically the most complex when dealing with RDSs since these images are free of any relevant information other than binocular disparity. In this section, we will focus on the wiring of simple and complex cells of the cat and monkey primary visual cortex.

4.1 Solving the correspondence problem with Marr's computational approach

Using Julesz's RDS as a case study, Marr and Poggio (1979; 1976) developed an algorithm capable of extracting depth from binocular disparity. The authors constrained matching solutions by applying the constraints based on the physical properties of the world. To account for the fact that "disparity varies smoothly almost everywhere", they introduced a *smoothness constraint* (or continuity rule). Because any point has a unique position in space, the *uniqueness constraint* states that "each item from each image may be assigned at most one disparity value". Finally, corresponding points must have similar brightness or colour (*compatibility constraint*). A recent physiological study (Samonds, Potetz, & Lee, 2009) demonstrated the existence of local competitive and distant cooperative interactions in the primary visual cortex of the macaque, via lateral connections. These authors suggested that local competition could be the neural substrate of the uniqueness rule while distant cooperation would favour the detection of similar disparities and therefore implement the continuity rule.

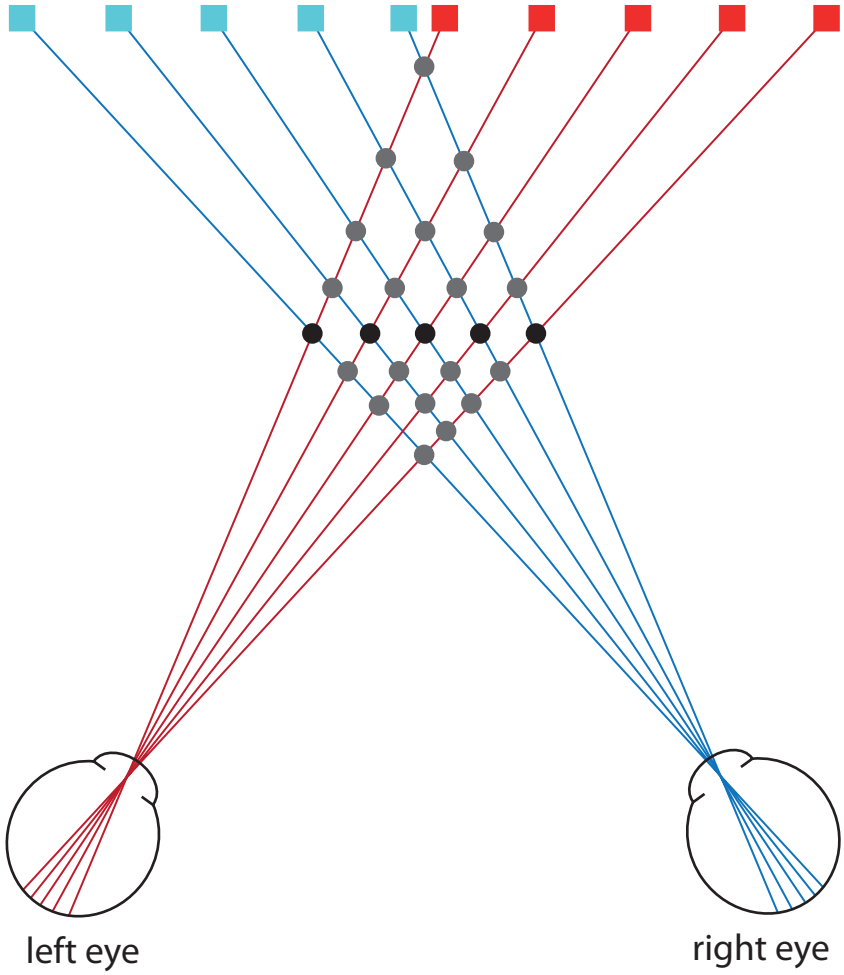


Figure III.9 |. Ambiguity in the correspondence between the left and right eyes projections. Each point in the left eye image could be matched with any of the points in the other image. All possible matches are shown in grey and black. Different rules based on ecological assumptions are used to constrain the algorithm into finding the most probable match (shown in black). (adapted from Marr & Poggio, 1976).

4.2 Position vs. phase disparity

Neurons in the visual cortex respond to stimulations in a defined region of the retina called the receptive field (RF). RFs of simple cells in primary visual areas can be described as a sinusoidal sensitivity function modulated by a Gaussian envelope (Fig. III.10). The size of the RF is represented by the variance of the Gaussian. The sensitivity profile is determined by a cosine function with given frequency and phase. A binocular simple cell responds preferentially to a grating of given frequencies and phases for the left and right eyes.

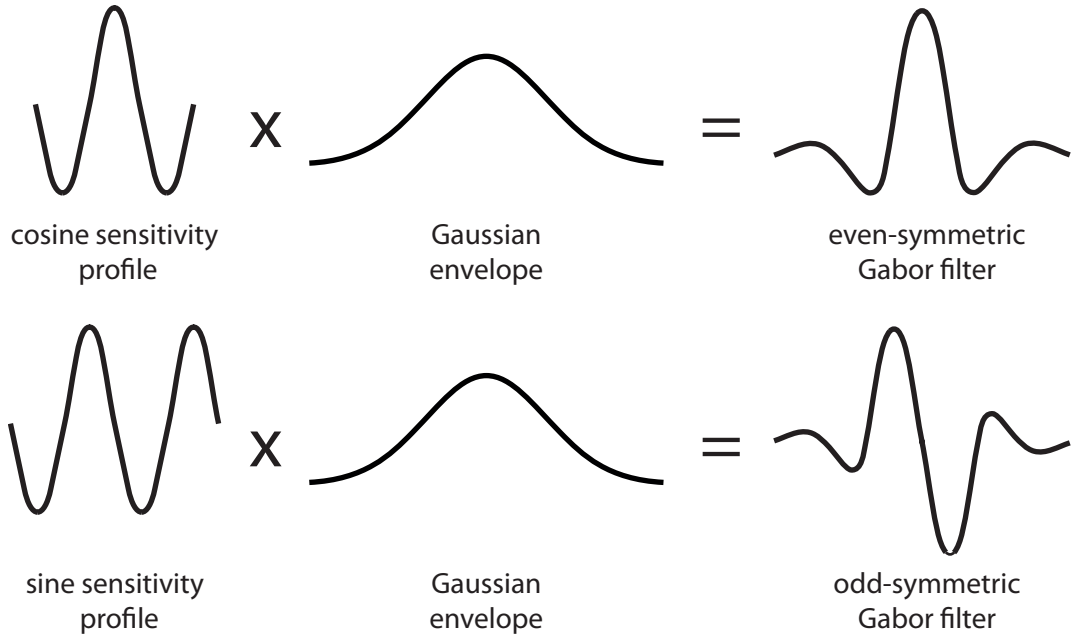


Figure III.10 | Sensitivity profiles of simple-cell receptive fields. The sensitivity profile is obtained by multiplying a carrier sinusoidal sensitivity function with a Gaussian envelope. Cosine carriers result in even-symmetric RFs and sine carriers result in odd-symmetric RFs. (adapted from Howard, 2002).

To detect disparities different from zero, the receptive fields in the two eyes must differ. Disparity detection can be achieved either by shifting the position of the RF (position disparity detectors) in one eye relative to the other or by shifting the phase of the cosine sensitivity profile in one eye relative to the other (phase disparity detectors).

In the case of position disparity detectors, the left and right eyes RFs feeding into the binocular simple cell have identical shapes and vary only in their horizontal position (a shift of the envelope). The shift in horizontal retinal position signals the disparity. In this type of disparity detectors, the spatial frequency of the RFs and the position shift are independent. A high spatial frequency RF can detect large disparities and vice versa. This mechanism allows the detection of substantially large disparities and, as a consequence, is prone to signal false matches.

In the case of phase disparity detectors, the left and right eyes RFs have identical sensitivity profiles but different distributions of excitatory and inhibitory zones (a shift of the carrier). The preferred disparity equals the phase shift divided by the spatial

frequency of the stimulus. In this type of mechanism, uncertainty is increased by the fact that the disparity measure depends on both the phase shift and the spatial frequency. It is hypothesized that this uncertainty is decreased by pooling over orientation and position (Tyler & Julesz 1980). Because the maximum detectable disparity is proportional to the spatial frequency in the RFs, small disparities are detected by high spatial frequency sensitive binoculars cells and large disparities by low spatial frequency cells.

Neurophysiological recordings have demonstrated the existence of these two types of disparity detectors (Prince, Cumming & Parker, 2002) and that many binocular simple cells show a combination of both phase and disparity shift (Tsao, Conway, & Livingstone, 2003).

It can be hypothesized that the two types of detectors carry out complementary processes. For example, position disparity detectors are not limited in size. Therefore, they could theoretically detect very large disparities and sustain depth perception in diplopic displays. On the other hand, phase disparity detectors could theoretically signal disparity between features of opposite polarity in the two eyes. This specificity could explain neurophysiological and psychophysical data such as the detection of anticorrelated stereograms by primary visual cortical neurons (Cumming & Parker, 1997; Masson, Busettini, & Miles, 1997) and double fusion as in the Panum's limiting case or da Vinci stereopsis (Gillam, Blackburn, & Cook, 1995) (see chapter IV).

4.3 Complex cells and the disparity energy model

Similarly to simple cells, complex cells show selectivity for particular visual attributes such as orientation or disparity. However, unlike simple cells, complex cells show a certain degree of spatial invariance. They exhibit large RFs and respond to the presence of the appropriate attribute within the receptive field, independent of its exact location or phase. Complex cells combine inputs from several simple cells and their activity results from the integration and summation of the activity of the simple cells in their own RFs.

To explain the pattern of activity of complex binocular cells in the cat's cortex, Ohzawa, deAngelis and Freeman (1997) proposed that these cells act as disparity energy detectors (Fig. III.11).

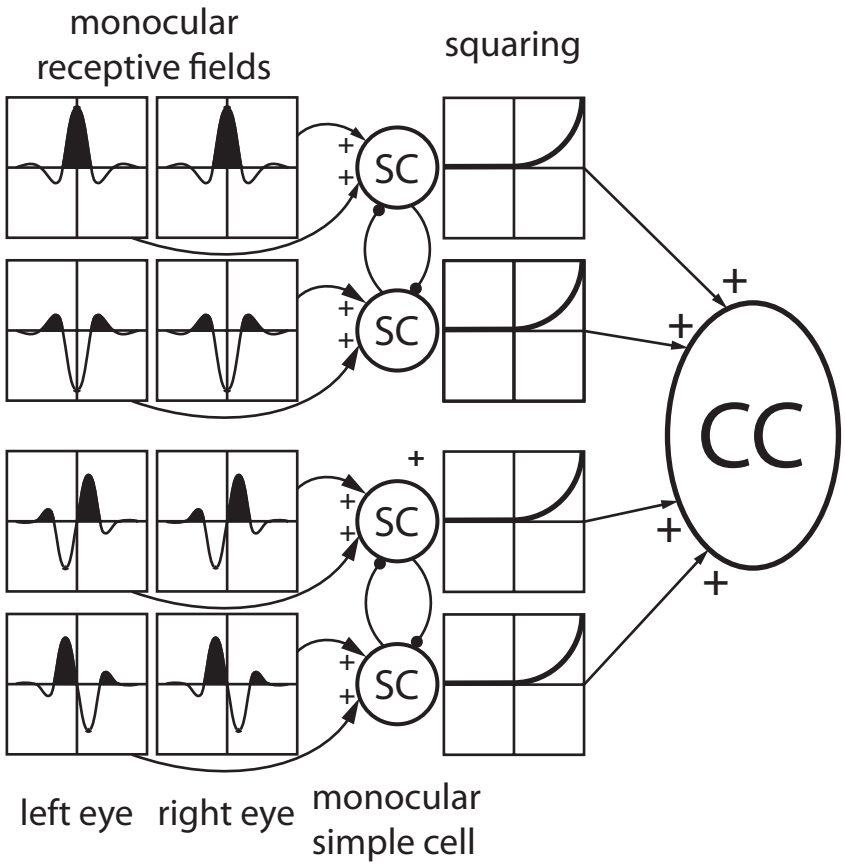


Figure III.11 | Illustration of the disparity energy model. Four binocular simple cells (SC) are combined by a complex cell (CC) tuned to zero disparity. Each simple cell receives inputs from the to eyes. The four subunits are arranged in mutually inhibitory pairs. The black and white areas represent excitatory and inhibitory regions respectively. (adapted from Howard, 2002).

A complex cell integrates the activation of four binocular simple cells that elicit different sensitivity profiles (phase dependence) but identical spatial frequency. The subunits are arranged in mutually inhibitory pairs, one in phase and one in quadrature phase (90°). Activations from the four subunits are squared and summed, resulting in an activation that is independent of the phase and position invariant in the RF of the complex cell. The preferred disparity of a complex cell is defined by the relative phase between left and right eyes RFs divided by the spatial frequency of the RF profiles of

the constituent subunits. When presented with anticorrelated stereograms, these complex cells show reversed disparity tuning functions (Cumming & Parker, 1997), supporting the validity of the disparity energy model.

4.4 Solving the correspondence problem with cross-correlation

The output from a bank of complex cells each tuned to a different disparity is then used to solve the correspondence problem, that is to say, eliminate false matches and construct a map of correct matches. A local cross-correlation mechanism is thought to be a good candidate for this job. Cormack, Stevenson & Schor (1991) were the first to mention that stereoacuity depends on the interocular correlation of the image intensity distributions.

To compute cross-correlation between the two images, two Gaussian correlation windows are moved independently in the two images (one vertically and one horizontally). A cross-correlation between the two windows is computed for each combination of window position for the two eyes. The output of the cross-correlator is a map of correlations as a function of the position of the Gaussian window in each eye. The correlation varies between -1 and +1 and the disparity pattern is revealed by peaks of high positive correlation (Banks et al., 2004). The main difficulty in implementing a cross-correlator algorithm is to determine the optimal size for the image patches sampled in each eye (Kanade & Okutomi, 1991). Patches that are too large may not be sensitive to small disparities while too small patches might no contain enough information to compute the correlation. Two studies found that the smallest useful mechanisms has a diameter of 3-6 arcmin (Filippini & Banks, 2009; Harris, McKee, & Smallman, 1997). Neurophysiological recordings and psychophysical data have provided evidence that cross-correlation mechanisms can reliably explain limitations of the stereoscopic system such as stereoresolution and the disparity-gradient limit (Banks et al., 2004; Filippini & Banks, 2009; Nienborg et al., 2004).

5 Conclusions

Over the past 20 years, our understanding of stereopsis has benefited from substantial advances in neurophysiology, imaging and modelling. The disparity energy model, developed by Ohzawa, deAngelis & Freeman (1997) explains a majority of the psychophysical and neurophysiological data collected until now. Moreover, the Maximum-Likelihood Estimation model (see chapter VI, section 1.4), proposed by Ernst & Banks (2002) to model multisensory integration has proven to be a good predictor of visual cue integration for the perception of depth (Ban, Preston, Meeson, & Welchman, 2012). However, several issues remain to be addressed. For example, more psychophysical and modelling studies are needed to better understand the respective role of position and phase disparity detectors. Up to now, imaging and single-unit recording studies have provided conflicting results on the processing of binocular disparity in the ventral and dorsal streams (Neri, Bridge, & Heeger, 2004; Preston, Kourtzi, & Welchman, 2009). Combining psychophysical and imaging methods might allow us to reconcile conflicting data collected up to now. Another issue is the integration of monocular occlusion cues and classic binocular disparity in the resolution of the correspondence problem. The next chapter (IV) presents a detailed review of this issue together with our first experimental study.

Part 2

Experimental Work

IV Depth perception from monocular occlusion

1 Introduction

1.1 History

In his book *Optics* (published about 300 BC), Euclid describes that the two eyes obtain different views of an object and that more of it can be seen with two eyes than one. Almost two millennia later, in 1508, Leonardo da Vinci noticed that next to a vertical edge of an opaque object is a region of a far surface that is visible to only one eye. In fact, when trying to picture a scene from the cyclopean view, he noticed that it is impossible to reproduce what is seen in three dimensions by the two eyes on a canvas.

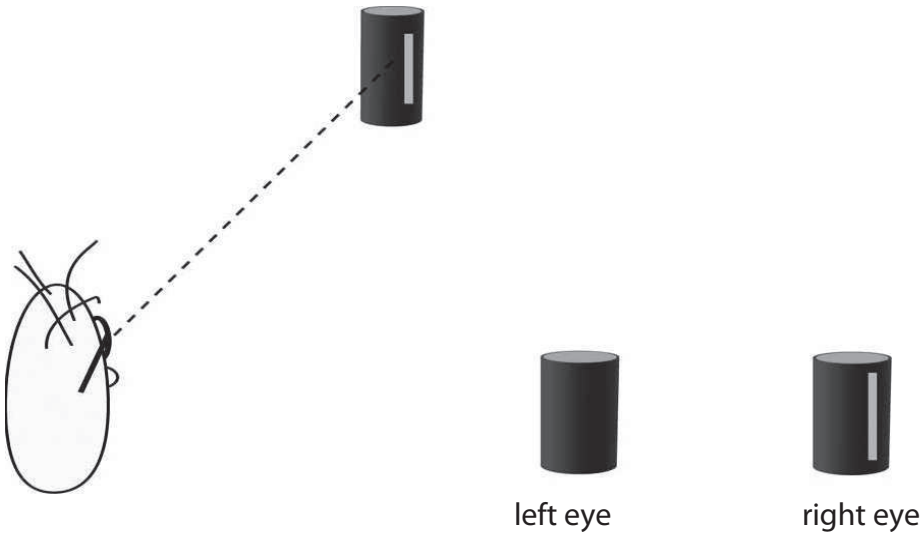


Figure IV.1 | "The phenomenon of binocular half-occlusion. The observer views an object (here a cylinder) binocularly. The light strip along the right portion of the object depicts a region visible only to the right eye, as shown in the images depicting the monocular views" (reproduced from Wilcox, 2007).

In a natural situation, if an object occludes a part of the visual scene, some areas of the configuration are seen by one eye only. There exist a lot of discontinuities due to the boundaries of objects. These abrupt changes in depth create a number of points that are present in one retinal image only (Fig. IV.1). One can assume that the visual system automatically ignores these monocular points to solve the correspondence problem. However, a majority of these unpaired points present in natural visual scenes carry crucial information about depth relationships between objects. Surprisingly, psychophysical, electrophysiological and computational studies did not recognize the potential influence of half-occlusion information on stereopsis and depth perception until the late sixties. The first study on the topic conducted by Lawson & Gulick (1967) demonstrated that occlusion cues can signal a depth offset. Twenty years later, Gillam & Borsting (1988) showed that it takes less time to detect a depth edge in a random dot stereogram (RDS) in the presence of half-occlusion regions that are congruent with the disparity information. To do so, these authors added patches of unpaired dots next to the left and right edges of a rectangle defined by binocular disparity. When the position of the unpaired regions was congruent with the geometry of occlusion (at the left of the rectangle in the left eye or at the right of the rectangle in the right eye — see Fig. IV.2 & IV.3) the detection of the depth edges was faster. Later, Anderson (1994) demonstrated that binocular features are actively decomposed into disparities and half-occlusions and that vertical image differences can signal occlusion and therefore generate a percept of depth. Research on monocular occlusion has mainly focused on two perceptual phenomena, namely *da Vinci stereopsis* (Nakayama & Shimojo, 1990) and *monocular gap stereopsis* (Gillam, Blackburn, & Nakayama, 1999). In both, parts of the visual scene that are present in one eye only are perceived accurately in depth even though there is no disparity information available to compute their location in space.

In the present review, we will first introduce da Vinci stereopsis and monocular gap stereopsis and explore whether these phenomena can be explained by classical stereoscopic mechanisms or whether they require the use of specific assumptions on the geometry of the scene. In a second part, we will present recent computational and

biologically inspired models of binocular processing that integrate unpaired features at varying levels of processing.

1.2 da Vinci stereopsis and occlusion geometry

1.2.1 Different types of monocular regions

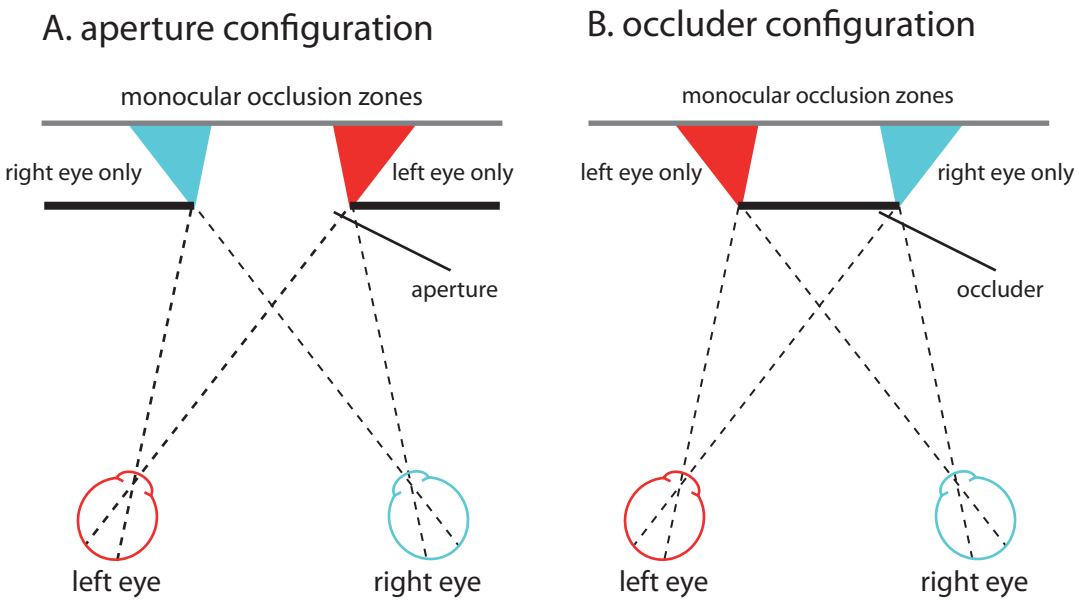


Figure IV.2 |Top view of two examples of geometrical configurations resulting in monocular regions. A. Aperture configuration: looking at a distant surface through a central square aperture. B. Occluder configuration: looking at a central square in front of a background. In both cases, specific regions of space are visible only to the left or the right eye.

In Figure IV.2a, the background is seen through an aperture that is smaller than the interocular distance. In Figure IV.2b, an object smaller than the interocular distance is seen binocularly. Different parts of the background are occluded to each eye. The difference in visual direction for the two eyes creates zones that can only be seen only by one eye.

1.2.2 Da Vinci stereopsis stimulus

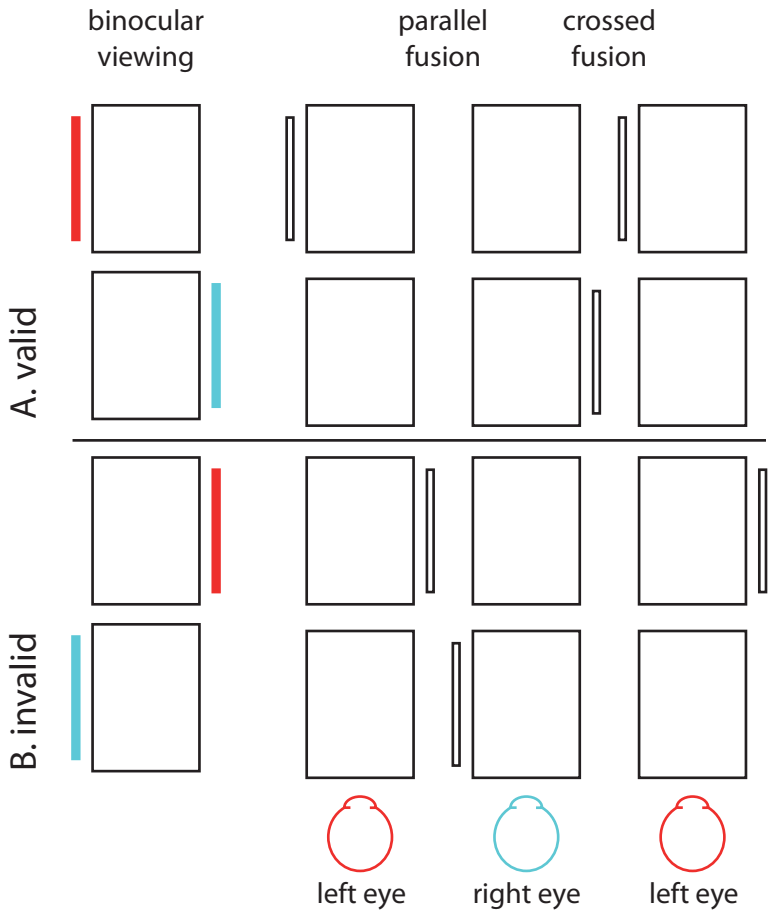


Figure IV.3 | Description of the conditions used by Nakayama & Shimojo (1990). The monocular line is presented close to a binocular rectangle. In the "valid condition", the line is presented in the temporal side of the rectangle: to the left in the left eye or to the right in the right eye. . The "invalid condition" is obtained by switching the two eye's views from the "valid condition": the line is presented in the nasal side of the rectangle: to the right in the left eye or to the left in the right eye.

On the basis of da Vinci's drawings, Nakayama & Shimojo (1990) used a simple stimulus configuration where a monocular vertical line is presented close to a binocular rectangle (Fig. IV.3) to investigate the role of the stimulus geometry and ecological validity on the perceived depth of monocular points. In this half-occlusion configuration, the rectangle acts as an occluder. When the line is presented in an ecologically *valid* configuration (on the temporal side of the occluder), the line is perceived at a precise depth that depends on the line-occluder distance (or line eccentricity). They called this impression of depth *da Vinci stereopsis*. On the contrary,

when presented to the nasal side (*invalid* condition), the line is perceived at the depth of the occluder.

1.2.3 Occlusion geometry

To explain their results, Nakayama & Shimojo (1990) postulated that the visual system is able to extract the geometry of the scene and the occlusion relations in it. Then, the position of the monocular objects, the eye-of-origin information and the geometry are combined to compute the perceived depth of the unpaired points. The edges of the occluder define constraint lines delimitating a constraint zone. This constraint zone hidden to one eye defines the area in which a monocular object must lie to refer to an ecologically valid situation (Fig. IV.4). As the eccentricity from the occluder increases, the corresponding monocular occlusion zone is displaced further in depth (Fig. IV.4). Therefore, in this valid condition, the perceived depth of a monocular object increases with eccentricity.

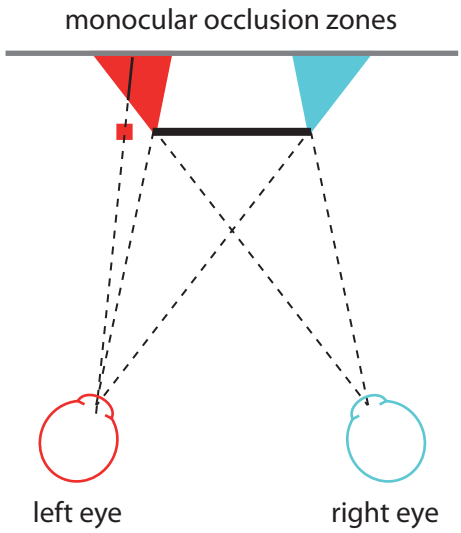


Figure IV.4 | Constraint lines and constraint zones. The constraint zone is defined by two constraint lines: one joining the eye to which the monocular line is presented and the line and another one joining the other eye and the occluder's edge adjacent to the monocular line. When presented in an ecologically valid condition, the monocular object is perceived along the eye-object constraint line, into this constraint zone (anywhere along the solid segment of the eye-object constraint line). In this drawing, the line (red square) is seen only by the left eye.

1.2.4 First data

Nakayama & Shimojo (1990) reported that for their stimuli, the perceived depth corresponded to the minimal possible depth (nearest constraint line) but did not provide a theoretical explanation for this observation. This minimal depth effect could possibly be accounted for by the fact that the visual system tends to minimize local differences in disparity when faced with an ambiguous visual scene (Goutcher & Mamassian, 2005).

Beyond an eccentricity of 30-40 arcmin, the line regresses to the occluder depth. The authors had no convincing explanation for this result either. Hakkinen & Nyman (1996) replicated Nakayama & Shimojo's observation of regression to the occluder plane (but beyond an eccentricity of 10-15 arcmin) and interpreted this result according to a capture constraint: beyond a given eccentricity, the depth of the monocular object is captured by the binocular elements present in the scene (here, the occluder). This result is also compatible with the bias for small disparities observed by Goutcher & Mamassian (2005).

It is worth mentioning that the "invalid condition" of Nakayama & Shimojo is actually a camouflage configuration. If the monocular object has the same texture and luminance as the foreground, it is "camouflaged" in one eye (and therefore invisible) and not in the other. Interestingly, according to Nakayama & Shimojo's results the visual system does not seem to treat occlusion and camouflage equally, considering camouflage as very unlikely (but see Cook and Gillam, 2004) for a case in which camouflage was easier than occlusion).

Ono, Wade & Lillakas (2002) and Ono, Lillakas, Grove, & Suzuki (2003) reformulated da Vinci and Nakayama & Shimojo's observations in terms of direction. Two opaque objects cannot be seen in the same direction. When the distance between the occluder and the occluded object is small, to satisfy this "Leonardo's constraint" the visual system compresses and shifts some elements of the visual scene that are located behind the fixated object. This way these elements are perceived next to the occluding object and not behind.

1.3 da Vinci stereopsis and double fusion

A few years after Nakayama & Shimojo's study, several authors pointed out the similarity between their da Vinci stimulus and the Panum's limiting case.

1.3.1 Panum's limiting case

In 1858, Panum described a natural situation in which two vertical lines at different depths are seen in a single direction for one eye, so that their images for that eye are superimposed, but lie in different directions for the other eye, resulting in two separate images (Fig. IV.5). In other words, when two vertical lines presented to one eye are fused with a single line presented to the other eye, they are perceived as two lines in depth (Hering, 1861; Panum, 1958). This depth effect can be explained by a double fusion process in which the single line is fused separately with each of the two lines in the other image (Gillam et al., 1995). The resulting depth depends on the disparity between the two lines. The Panum's limiting case violates the uniqueness constraint stated by Marr & Poggio (1976): "each item from each image may be assigned at most one disparity value"

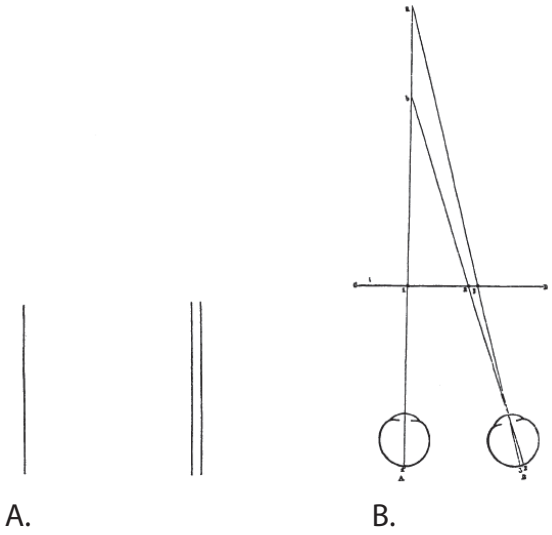


Figure IV.5 | Panum's limiting case. A. The Panum stereogram: the single line presented to the left eye is fused with both lines presented to the right eye: the right line appears further away (positive disparity). B. A configuration that could give rise to the Panum's limiting case: the images of the two lines are superimposed in one eye's (left) view but not in the other (right). (reproduced from Panum, 1858)

1.3.2 Da Vinci stereopsis explained by double fusion

Due to similarities between the Panum's limiting case and da Vinci stereopsis, some authors have tried to find a common explanation, proposing that one is a simple variation of the other. Ono, Shimono & Shibuta (1992) reported results similar to Nakayama & Shimojo's findings with a Panum's limiting case stimulus and hypothesized that it is a special case of da Vinci stereopsis. Gillam, Blackburn and Cook (1995) used a stimulus similar to Ono et al. (1992) but controlled for vergence eye movements and line eccentricity and obtained results favouring a double fusion explanation for both Panum's limiting case and da Vinci stereopsis. In other words, according to Ono et al. (1992) and to Gillam et al. (1995), the adjacent edge of the occluder in one eye's image would be "double-fused" with its counterpart and the monocular line in the other eye's image. The line would be seen in front or behind the occluder depending on the eye to which the line is presented. Later, Gillam, Cook & Blackburn (2003) designed a da Vinci stimulus in which the monocular object is a disk that cannot be "double-fused" with the adjacent edge of the occluder. They found that the depth perception of the disk was qualitative: it was always perceived as lying behind the occluder and the occluder-disk separation had no effect on the perceived depth. These authors concluded that fusibility is a critical factor for seeing precise quantitative depth, confirming that Nakayama & Shimojo's results can be explained by double-fusion.

1.3.3 Issues pending

Even though the experiments reported in the previous paragraph support the idea that the *quantitative* depth percepts observed in Nakayama & Shimojo's study (1990) might be due to double matching, other aspects of their results cannot be accounted for by standard stereoscopic mechanisms.

For example, Nakayama & Shimojo (1990) and Häkkinen & Nyman's (1996) finding that the perceived depth of the monocular line regresses to the occluder's plane for eccentricities larger than 30-40 arcmin and 10-15 arcmin respectively is incompatible with the properties of the Panum's fusional area. Studies on the spatial limitations of stereopsis have reported that disparities up to 125 arcmin can elicit a

reliable percept of depth (Schor & Tyler, 1981). In addition, it has been shown that diplopic stimuli still elicit a qualitative percept of depth (Wilcox & Allison, 2009). If da Vinci stereopsis is resolved by double matching (i.e. through conventional stereopsis mechanisms), eccentricities beyond 30-40 arcmin should be treated accurately.

Gillam, Cook & Blackburn's (2003) monocular disk was systematically perceived behind the occluder's plane and the authors did not provide an explanation for this observation. This observation suggests that in the absence of disparity information, monocular objects are positioned behind the occluder's plane by default.

To address these various pending issues, we conducted two experiments and derived a simple model to explain our data. This work is presented in the form of a published article in section 2 of this chapter.

1.4 Monocular gap stereopsis

In 1999, Gillam, Blackburn & Nakayama (1999) designed a novel configuration in which the perceived depth could not be accounted for exclusively by classic stereopsis mechanisms. In the so-called *monocular gap stereopsis*, one eye sees one black rectangle and the other the same rectangle with a central gap. The resulting percept consists of two flat rectangles seen at different depths (Fig. IV.6a).

The right eye's view is obtained by introducing a central gap in the left eye's image. The addition of this empty white region creates disparities at the outer edges of the entire fused configuration. Based on classic stereoscopic mechanisms, one would predict that this stimulus would be perceived as a slanted plane with a rivalrous central patch at the location of the monocular gap (Fig. IV.6). However, based on ecological geometry of occlusion, the occurrence of such a monocular gap is only coherent with the existence of two flat surfaces separated in depth. Therefore, monocular gap stereopsis appears to be a pure example of depth from occlusion. As shown for da Vinci stereopsis, Gillam et al. (1999) observed that the perceived depth between the two surfaces increases with the size of the gap.

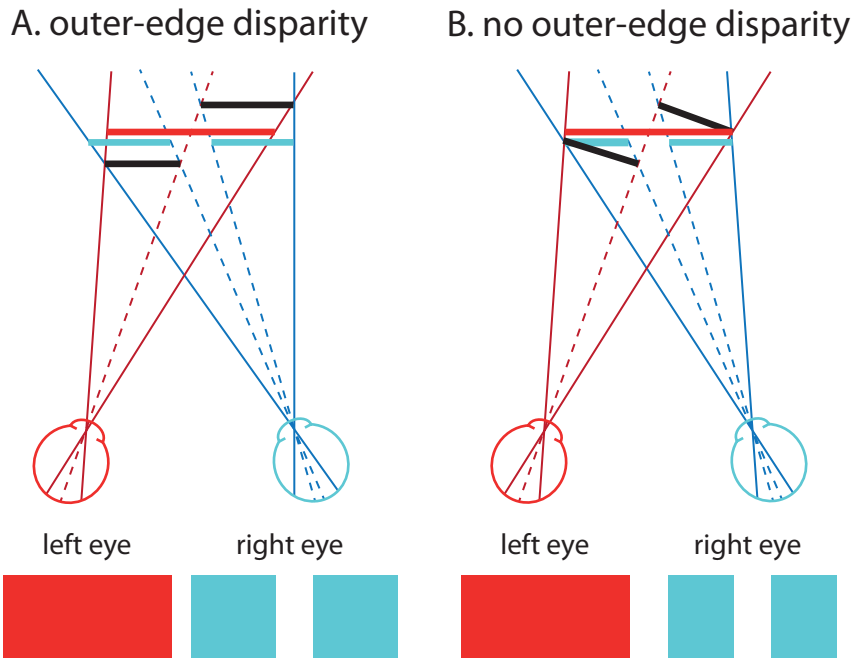


Figure IV.6 | Monocular gap stereopsis. The right eye can see a white background through the gap between the two objects. This gap is occluded to the left eye by the foremost object. Red and cyan bars represent the views from the left and right eye respectively. Solid lines represent the lines of sight for the stimuli and dashed lines represent the central partitioning (theoretical in the case of the left eye's view). Black bars represent the percepts predicted by classic stereoscopic mechanisms using the aforementioned partitioning. (adapted from Pianta & Gillam, 2003b).

To investigate the mechanisms underlying monocular gap stereopsis, Pianta & Gillam (2003a) compared monocular gap stimuli and binocular gap stimuli (a central gap is present in the two eyes' images) and found identical thresholds for the two. More interestingly, they found that adapting to a binocular gap led to shifts in the perceived depth of monocular gap stimuli and vice versa. These two observations led these authors to suggest that monocular gap stereopsis is processed by classic stereopsis mechanisms. However, it is worth mentioning that the cross-adaptation found in their study might take place at a higher level of processing, after monocular regions and classic stereopsis are processed by two separate mechanisms. In a follow-up study, Pianta & Gillam (2003b) manipulated the disparity of the outer edge of the solid rectangle (Fig IV.6a & IV.6b). When outer-edge disparities are present in the stimulus, the left eye sees one solid black surface while the right eye is presented with the same object partitioned and presented with a central gap in between. The addition

of this gap yields to the presence of outer-edge disparities. In this configuration, two solid flat objects are seen at different depths. In order to remove outer-edge disparities, the right eye image is shrunk in width so that the total width is equal to the width in the left eye. This configuration yields to the perception of two solid slanted planes seen with a maximum depth different at the centre.

They measured depth thresholds with and without outer edge disparity and found that depth was perceived at the gap even when the two images had the same width (no outer edge disparity) and that this depth varied with the size of the gap. This result provided even stronger evidence that monocular gap stereopsis is mediated by non-classic stereoscopic mechanisms. To test the importance of geometry in monocular gap stereopsis, Grove, Sachtler & Gillam (2006) added two black squares at the end of the gap of a grey monocular gap stimulus. They showed that the perceived depth of the gap was attenuated when the two black squares were placed stereoscopically behind the monocular gap configuration but not in front. These authors argued that amodal completion between the gap and the background is necessary in monocular gap stereopsis. Therefore, placing two black squares behind the configuration strongly disturbed this amodal completion, suggesting a critical implication of geometry in monocular gap stereopsis.

To complement these geometrical manipulations, Grove, Gillam & Ono (2002) manipulated the textures of the background and monocular gap and found that the perceived depth at the location of the gap was dramatically impaired when the background and gap textures did not match.

1.5 Stereo models including unpaired features

Classical models of stereo matching treat unpairable features as noise (Marr & Poggio, 1979). However, as cited above, several authors have reported a collection of evidence showing that monocular regions can convey reliable information about geometrical configuration and depth orderings. There are two possible approaches to integrate depth cues from unpaired features.

1.5.1 Late integration of monocular regions to the depth map

Monocular regions can be included during the final stages of stereo matching, to refine the disparity map (Jones & Malik, 1992): this map is processed post-hoc to determine the likely localizations of depth discontinuities. In this view, occlusion relationships must be derived from the geometry of the scene before they can be integrated to the depth map. Unpaired features thus cannot be used to facilitate the construction of stereoscopic depth.

1.5.2 Early detection of monocular regions by disparity detectors

Another option is to postulate that there exist early mechanisms capable of detecting monocular regions and occluding contours. In this view, occlusion geometry can serve as a depth cue to constrain the resolution of the matching problem (by excluding unpaired points as matching candidates) and construct the depth map of the scene. Anderson & Nakayama (B. L. Anderson & Nakayama, 1994) demonstrated that half-occlusions can bias the interpretation of an ambiguous stereoscopic pattern as soon as stereoscopic depth is resolvable, showing that occlusion geometry can impact the early stages of disparity processing. Since the middle nineties, different types of early-extraction models have been proposed.

Grossberg & Howe (2003) proposed a model of 3D surface reconstruction in which the lateral geniculate nucleus, V1, V3 & V4 use both monocular and binocular information to extract boundary representations and construct a depth map of the scene. Based on the Bayesian approach, Geiger, Ladendorf & Yuille (1995) described a model using the constant relationship in which a depth discontinuity in one eye always corresponds to an interocularly unpaired region in the other eye.

In the same vein, Watanabe & Fukushima (1999) developed a two-step stereo algorithm based on an occlusion constraint: an occluding point should exist between an unpaired point and the eye that cannot see the unpaired point. First, matching primitives are classified as paired or unpaired and eye-of-origin information is extracted. Then, these three types of data are combined to create the depth map. Hayashi, Maeda, Shimojo & Tachi (2004) extended Watanabe & Fukushima's model (1999). Using a classical disparity energy model, monocular regions are detected by

monitoring the output of a population of binocular neurons. When there is no consistent disparity signal (i.e. when features are present in only one eye), binocular neurons elicit a broad activation across a large band of disparity values. This specific pattern of activation is used to signal the presence of monocular regions. In addition, they proposed that the detection of unpaired features could be achieved by an interocular inhibition mechanism since it is contradictory for monocular regions to be present in both eyes. This additional occlusion constraint provides an interesting model of binocular rivalry. When two monocular regions are present in the same location, their mutual interocular inhibition results in an unstable output that alternates between the two possible interpretations. This model is the first to integrate disparity processing with monocular regions and binocular rivalry.

Assee & Qian (2007) pointed out the fact that these models are not parsimonious and that some of them postulate the existence of specific monocular cells. Against this, they proposed a model based on a simple V1-V2 feedforward structure. Depth edges and monocular regions are extracted in V2 from the outputs of V1 binocular cells.

Based on existing knowledge about the physiology of stereopsis, Tsao, Conway & Livingstone (2003) proposed that half-occlusions can be signalled by using a combination of phase and position shifts, giving an ecological justification for the existence of these two types of coding.

1.6 Conclusion

While there has been a vigorous debate on whether da Vinci stereopsis is processed by classic stereo mechanisms or using occlusion geometry, there is a consensus around the idea that monocular gap stereopsis cannot be fully accounted for by classic stereoscopic mechanisms.

In the experimental work presented in the following section, we address whether da Vinci stereopsis is processed by classic stereopsis or using occlusion geometry. To do so, we used a simple configuration (Nakayama & Shimojo, 1990) and manipulated the material properties of the occluding object.

2 The role of transparency in da Vinci stereopsis

Vision Research 51 (2011) 2186–2197

Contents lists available at SciVerse ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

The role of transparency in da Vinci stereopsis

Marina Zannoli *, Pascal Mamassian

Laboratoire Psychologie de la Perception (CNRS UMR 8158), Université Paris Descartes, France

ARTICLE INFO

Article history:
Received 7 February 2011
Received in revised form 12 August 2011
Available online 28 August 2011

Keywords:
da Vinci stereopsis
Half occlusion
Perceptual transparency
Binocular models

ABSTRACT

The majority of natural scenes contains zones that are visible to one eye only. Past studies have shown that these monocular regions can be seen at a precise depth even though there are no binocular disparities that uniquely constrain their locations in depth. In the so-called da Vinci stereopsis configuration, the monocular region is a vertical line placed next to a binocular rectangular occluder. The opacity of the occluder has been mentioned to be a necessary condition to obtain da Vinci stereopsis. However, this opacity constraint has never been empirically tested. In the present study, we tested whether da Vinci stereopsis and perceptual transparency can interact using a classical da Vinci configuration in which the opacity of the occluder varied. We used two different monocular objects: a line and a disk. We found no effect of the opacity of the occluder on the perceived depth of the monocular object. A careful analysis of the distribution of perceived depth revealed that the monocular object was perceived at a depth that increased with the distance between the object and the occluder. The analysis of the skewness of the distributions was not consistent with a double fusion explanation, favoring an implication of occlusion geometry in da Vinci stereopsis. A simple model that includes the geometry of the scene could account for the results. In summary, the mechanism responsible to locate monocular regions in depth is not sensitive to the material properties of objects, suggesting that da Vinci stereopsis is solved at relatively early stages of disparity processing.

1. Introduction

There is more to binocular vision than the matching of corresponding objects in the left and right images. Since the early physiological recordings of Hubel and Wiesel in cats (1959), binocular disparity was thought to be processed in area V1 and extrastriate areas (MT in primates) primarily (Howard & Rogers, 2002; Parker, 2007). Within the last decade this classical view has been challenged by several studies in electrophysiology and imaging indicating that disparity processing might be distributed across several regions of the visual cortex (Backus et al., 2001). For example, Preston et al. (2008) showed that areas V3 and V4 are sensitive to both correlated and anticorrelated stimuli. These results suggest that there exist many steps of processing between the extraction of the disparity signal to the computation of the depth map. One of them consists in determining depth ordering relationships between objects, namely which object is in front of another without any precise estimate of the distance between the two. Traditionally, depth ordering has been associated with monocular cues based on luminance such as transparency (Anderson, 2008) or occlusion (Sekuler & Palmer, 1992). Yet, binocular cues can be equally efficient in conveying depth ordering information. In particular, da Vinci stereopsis provides a convincing illustration of the interaction between occlusion and stereopsis.

1.1. da Vinci stereopsis and occlusion geometry

In 1508, Leonardo da Vinci noticed that next to a vertical edge of an opaque object is a region of a far surface that is visible to only one eye (see Fig. 1). Boundaries of objects produce a lot of depth discontinuities. These abrupt changes in depth can create a number of points that are present in one retinal image only. One can assume that the visual system automatically ignores these monocular points to solve the correspondence problem. However, a majority of these unpaired points present in natural visual scenes carry crucial information about depth relationships between objects (see Harris and Wilcox (2009) for a comprehensive review). The first study on the role of half-occlusions, conducted by Lawson and Gulick (1967), demonstrated that occlusion cues can signal a depth offset. Later, Gillam and Borsting (1988) used random-dot stereograms and added half-occlusion regions that could be either congruent or incongruent with the disparity information. They showed that observers were faster to detect a depth edge in the congruent condition than in the incongruent case. Two types of configurations can lead to the presence of monocular regions: occlusion and camouflage (see Fig. 1a).

* Corresponding author.
E-mail address: marinazannoli@gmail.com (M. Zannoli).

0042-6989/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.visres.2011.08.014

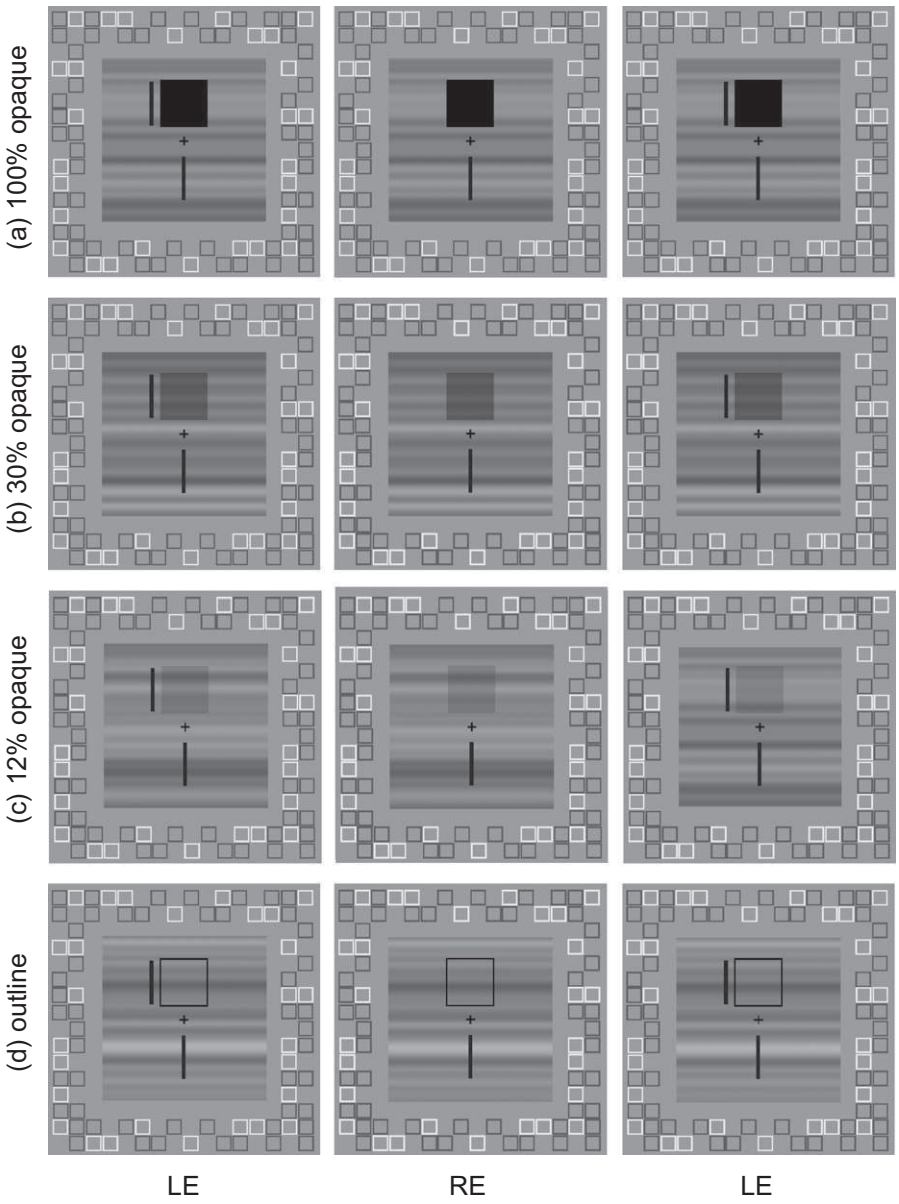


Fig. 1. Stimulus used in Experiment 1. The valid condition can be seen by parallel-fusing the first and second columns. The invalid condition is seen when parallel-fusing the second and third columns. (a) Classical da Vinci configuration where the occluder is completely opaque. (b) Condition where the occluder is 30% opaque. (c) Condition where the occluder is 12% opaque. (d) Condition where the occluder is just represented by its outline.

On the basis of da Vinci's drawings, Nakayama and Shimojo (1990) used a simple stimulus configuration where a monocular vertical line is presented close to a binocular rectangle to investigate the role of the stimulus geometry and ecological validity on the perceived depth of monocular points (see Fig. 1). In this half-occlusion configuration, the rectangle acts as an occluder. When the line was presented on the temporal side of the occluder (in an ecologically "valid" configuration), the authors found that the line was perceived at a precise depth that depended on the line-occluder distance (or line eccentricity). They called this impression of depth "da Vinci stereopsis". On the contrary, when presented to the nasal side ("invalid" condition), the line was perceived at the depth of the occluder (see Fig. 2 for detailed predictions). To explain these results, the authors postulated that the visual system is able to extract the geometry of the scene and the occlusion relations in it. Then, the position of the monocular objects, the eye-of-origin information and the geometry are combined to compute the perceived depth of the unpaired points. The edges of the occluder define constraint lines delimitating a constraint zone. This constraint zone hidden to one eye defines the area in which a monocular object must lie to refer to an ecologically valid situation. The perceived depth increases with eccentricity and corresponds to the minimal possible depth, defined by the nearest constraint line (Nakayama & Shimojo, 1990). Beyond an eccentricity of 30–40 arcmin, the line regresses to the occluder depth (Nakayama & Shimojo, 1990). Nakayama and Shimojo's "invalid condition" is obtained by switching the two eye's views from the "valid condition". In this case, if the monocular object has the same texture and luminance as the foreground, it is "camouflaged" in one eye (and therefore invisible) and not in the other. Interestingly, the visual system does not seem to treat occlusion and camouflage equally, considering camouflage as very unlikely (but see Cook and Gillam (2004) for a case in which camouflage was easier than occlusion).

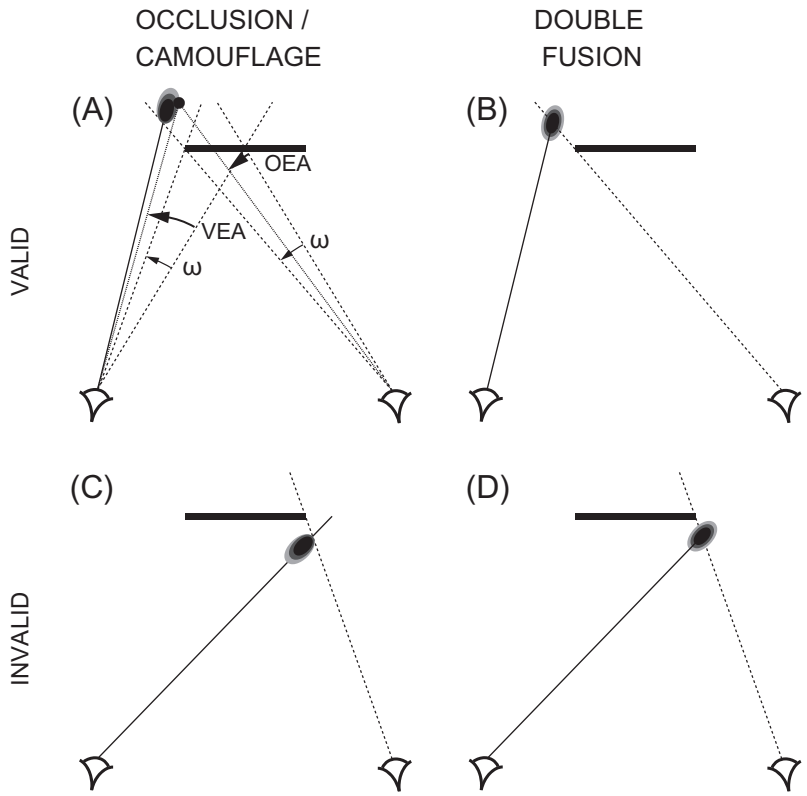


Fig. 2. Definitions of angles and predictions of the occlusion/camouflage and double fusion hypotheses. By convention, the monocular object is always presented to the left eye. (a) Definitions: the dot is an example of the location of the perceived monocular object for one trial, the *Other Eye Angle* (OEA) is its perceived depth for that trial and the *Viewing Eye Angle* (VEA) is its perceived azimuth. The ω angle represents half of the occluder's width. This figure also shows the predictions for the valid condition under the occlusion scenario: the predicted shape of the distribution of percepts is illustrated by contour plots (darker is more likely). (b–d) Predictions for the valid/double fusion case, the invalid/occlusion case and the invalid/double fusion case respectively.

1.2. Da Vinci stereopsis and double fusion

A few years later, several authors pointed out the similarity between the configuration used by Nakayama and Shimojo and Panum's limiting case. When two vertical lines presented to one eye are fused with a single line presented to the other eye, they are perceived as two lines in depth (Panum, 1858). This depth effect can be explained by a double fusion process in which the single line is fused separately with each of the two lines in the other image (Gillam, Blackburn, & Cook, 1995). The resulting depth depends on the disparity between the two lines.

Due to similarities between the two configurations, some authors have tried to find a common explanation, supposing that one is a simple variation of the other. Ono, Shimono, and Shibuta (1992) reported results similar to Nakayama and Shimojo's findings with a Panum's limiting case stimulus and hypothesized that it is a special case of da Vinci stereopsis. Gillam, Blackburn, and Cook (1995) used a stimulus similar to Ono, Shimono, and Shibuta (1992) and obtained results favoring a double fusion explanation for both Panum's limiting case and da Vinci stereopsis. In the latter case, the monocular line would be "double-fused" with the adjacent edge of the occluder in the other eye. The line would be seen in front or behind the occluder depending on the eye to which the line is presented (see Fig. 2 for detailed predictions). Later, Gillam, Cook, and Blackburn (2003) designed a da Vinci stimulus in which the monocular object is a disk that cannot be "double-fused" with the adjacent edge of the occluder. They found that the perceived depth was qualitative but not quantitative in the sense that it only signaled depth ordering. They also reported that this perceived depth depended on the validity of the scene configuration, suggesting a double fusion explanation for da Vinci stereopsis.

1.3. Aims of the study

The main aim of the present study was to investigate the importance of opacity on da Vinci stereopsis using perceptual transparency (Metelli, 1985; Singh & Anderson, 2002). If the degree of transmittance of the occluder influences the perceived depth in da Vinci stereopsis, this suggests that sophisticated aspects of the scene are taken into account during construction of the depth map as suggested by Nakayama and Shimojo. In contrast, if the processing of monocular regions does not depend on the opacity of the occluder, then low-level binocular mechanisms, such as double fusion, might be sufficient to explain da Vinci stereopsis. A secondary aim of the study was to estimate the consistency of the depth reports in da Vinci configurations. This consistency was measured by recording the whole distribution of depth percepts and by analyzing the spread and other statistical aspects of this distribution.

2. Experiment 1

To test whether da Vinci stereopsis is sensitive to the material properties of occluding objects, we manipulated perceptual transparency. According to the model of Singh and Anderson (2002), the opacity of a transparent surface is determined by the contrast ratio of the lower contrast regions (region of transparency) relative to the higher contrast regions (background) (see Fig. 1). We consider that this type of transparency has several advantages. First, the degree of opacity can be manipulated extremely precisely, allowing us to test whether opacity is fully required and whether it has a quantitative effect on da Vinci stereopsis. Psychophysical and neurophysiological studies suggest that the computation needed

to extract the transmittance (and thus the depth ordering) requires an intermediate level of processing (Qiu & von der Heydt, 2007; Singh & Anderson, 2002). Perceptual transparency thus represents a complex depth cue. Using such a mid-level cue allows us to assess the level of processing required to compute the occlusion geometry in da Vinci stereopsis.

2.1. Methods

2.1.1. Participants

Four naïve observers with normal or corrected-to-normal vision were recruited in the laboratory building. All participants had experience in psychophysical observation and had normal stereo acuity and transparency sensitivity.

2.1.2. Stimulus presentation

The stereograms were presented on a CRT monitor (ViewSonic 21″, resolution of 1280 × 960, refresh rate of 85.0 Hz) using a modified Wheatstone stereoscope at a simulated distance of 1 m. Each eye viewed one horizontal half of the CRT screen. A chin rest was used to stabilize the observer's head and to control the viewing distance. The monitor was linearized in luminance (gamma corrected). The display was the only source of light and the stereoscope was calibrated geometrically to account for each participant's interocular distance.

2.1.3. Stimuli

Stimuli were generated using the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). A binocular black (5 cd/m^2) square was presented in the upper visual field (1.3° from the center). We denote by ω the half width of the occluder: ω = 0.8°. A monocular black line of 0.1 × 1.6 deg^2 was presented next to the square. Another black line of 0.1 × 1.6 deg^2 was presented binocularly in the lower visual field. These three elements were presented on a textured background. The background was a 1-dimensional noise texture produced by blurring a texture of random 1-pixel-wide horizontal stripes with a vertical Gaussian (SD 1.15°). The background was comparable to a wallpaper stimulus, in the sense that there was a complete ambiguity on correspondence (see Fig. 1). The degree of opacity of the black square varied randomly between three values (100%, 30% and 12% opaque) chosen on the basis of pilot experiments. The transparent square was defined by changing the alpha index (Brainard, 1997; Pelli, 1997; Porter & Duff, 1984) of the binocular square region of the background area. An "outline" condition in which the binocular square was only defined by its edges (thickness of 0.03°) was added.

The distance between the monocular line and the black square varied randomly between three values. We denote by ε the eccentricity between the monocular line and the closest edge of the occluder. Three values were chosen for ε: 10, 19 and 28 arcmin. These values were chosen to match Nakayama and Shimojo's (1990) stimulus configurations. The eye of presentation (left or right) of the monocular line was counterbalanced and the side of presentation (left or right of the square) varied randomly to create four different conditions. In the "valid" condition, the line was presented to the temporal side of the square and in the "invalid" condition the line was presented to the nasal side (see Fig. 2).

The textured background was surrounded by a vergence-stabilization frame consisting of multiple black and white small squares (0.35 × 0.35 deg^2; black: 5 cd/m^2 and white: 80 cd/m^2) presented on a gray background (55 cd/m^2). Black nonius lines were added at the center.

2.1.4. Procedure

While keeping the nonius lines aligned, participants were asked to evaluate the perceived azimuth and depth positions of the monocular line using an adjustment procedure. The observers controlled the horizontal position and depth coordinates of the stereo-probe located in the lower visual field using the four keyboard directional arrows: the left and right arrows controlled for the azimuth position of the stereo-probe while the up and down arrow keys controlled for the depth. The stereo-probe appeared at the central position at the beginning of each trial. The impression of depth was created by adding positive or negative disparity to the lines between the two eyes' images. The participants were instructed to privilege accuracy rather than speed. Final spatial coordinates of the stereo-probe were recorded separately for the right and left image for each trial. Each combination of eccentricity values, eye-of-origin, opacity values and validity configurations was repeated 12 times in total. The experiment was divided in four sessions.

2.1.5. Data analysis

We define two visual angles to analyze the results. The *Viewing Eye Angle* (VEA) is the angle between the center of the occluder and the position of the probe for the eye that sees the monocular line. It gives an estimation of the horizontal position of the probe (*i.e.* the perceived azimuth of the monocular line – Fig. 2a). The *Other Eye Angle* (OEA) is the angle between the center of the occluder and the position of the probe for the eye that *does not* see the monocular line. It gives an estimation of the depth position of the probe (*i.e.* the perceived depth of the monocular line – Fig. 2a).

Data were pooled across the "side of the line" factor to bring the total number of trials per condition to 24.

2.1.6. Predictions

Different predictions can be advanced depending on the underlying explanations of da Vinci stereopsis.

2.1.6.1. Occlusion/camouflage hypothesis. If we follow strictly the occlusion geometry we predict that, in the valid condition, the monocular line should be occluded to the other eye and thus be perceived inside the far monocular zone (OEA < ω; see Fig. 2a). In the invalid condition, we predict that the monocular line would be camouflaged by the occluder to the other eye and therefore be perceived into the near monocular zone (*i.e.* again OEA < ω).

Extrapolating Nakayama and Shimojo's findings (1990), we can make slightly different predictions. We expect that the monocular line would be perceived on the near edge of the monocular zone (*i.e.* at the minimum possible depth: OEA ≈ ω) in the valid condition. In the invalid condition, we expect that the monocular line will be perceived at the depth of the occlusion plane (in this case, the fixation plane: OEA ≈ $\omega + \varepsilon$).

If da Vinci stereopsis relies on occlusion characteristics, we expect an effect of the opacity of the occluder on the perceived depth of the monocular line. More precisely, the impression of depth should decay as the occluder gets more transparent. In the extreme outline condition, perceived depth should be consistent with double fusion.

Regarding the perceived position of the line for the viewing eye, we naturally predict that its location should be veridical in both "valid" and "invalid" conditions (VEA ≈ $\omega + \varepsilon$; see Fig. 2a and c).

2.1.6.2. Double fusion hypothesis. According to the double fusion hypothesis, the distance between the monocular line and one edge of the occluder is processed as disparity. In this case, the line is seen in front or behind the occluder depending on the "validity" variable. This variable determines the sign of the disparity value. Following the double fusion hypothesis, we therefore expect that the monocular line would be perceived at the intersection of the line of sight going from the viewing eye to the monocular line and the line of sight going from the other eye to the adjacent edge of the occluder. Therefore, we expect the OEA and VEA coordinates

to be the same in both validity conditions (OEA $\approx \omega$ and VEA $\approx \omega + \varepsilon$; see Fig. 2b and d).

If da Vinci stereopsis is based on double fusion, we expect the opacity of the occluder to have no effect on the perceived depth of the monocular line.

2.1.6.3. Disentangling between occlusion and double-fusion. To sum up, occlusion and double fusion hypotheses give roughly the same predictions even though they rely on different underlying mechanisms. To disentangle the two explanations, we introduce a novel analysis using the shape of the distributions of depth estimations. In the double fusion hypothesis, OEA is treated as a disparity value whereas it represents a constraint line in the occlusion hypothesis. To account for this, we postulate that the distributions of perceived depths should be symmetrically distributed around the predicted value in the double fusion case: the uncertainty is equivalent in all depth directions. In contrast, in the occlusion case, we expect the distributions of perceived depths to be skewed to account for the constraints that define the monocular zones: the monocular line can be seen anywhere in the monocular zone but not outside this area (see Fig. 2a).

If surface material plays a role in da Vinci stereopsis, we expect a change in the skewness of the distributions of perceived depth with transparency in the occlusion case. A more opaque surface could more easily hide an object to the other eye, so there should be more skewness with more opacity.

2.2. Results

We treat the outline condition as a 0% opacity condition. Because no significant difference was found between the side of presentation conditions (left or right), OEA and VEA values were pooled across this factor and all results are presented as if they resulted from the left eye condition. When the monocular line is viewed by the left eye, it is presented on the left side of the occluder in the valid condition and on the right side in the invalid condition. The distributions of OEA and VEA reports are shown in Figs. 3 and 4.

2.2.1. Main effects of experimental variables

The OEA (depth) and VEA (azimuth) distributions were very consistent across subjects. Before conducting inferential analyses, we tested the normality of the OEA and VEA distributions obtained for each (eccentricity $*$ validity $*$ opacity) condition using the D'Agostino's normality test (D'Agostino, Belanger, & D'Agostino, 1990). Except for one VEA distribution ($\varepsilon = 19$ in the valid condition), all distributions were non-normal ($X2$ values ranging from 19.1 to 159). To take into account this non-normality, a repeated measures Analysis of Variance was conducted on the medians (and not the mean) for each validity condition separately. The ANOVA conducted on the OEA measures revealed a significant effect of eccentricity ($F(2,6) = 405$, $P < 0.001$ for the valid condition and $F(2,6) = 170$, $P < 0.001$ for the invalid condition) but no effect of opacity ($F(3,9) = 0.573$, $P = 0.647$ for the valid condition and $F(3,9) = 2.87$, $P = 0.096$ for the invalid condition – see Fig. 3). The ANOVA conducted on the VEA measures revealed the same pattern of results (eccentricity: $F(2,6) = 150$, $P < 0.001$ for the valid condition and $F(2,6) = 545$, $P < 0.001$ for the invalid condition; opacity: $F(3,9) = 3.24$, $P = 0.075$ for the valid condition and $F(3,9) = 0.426$, $P = 0.739$ for the invalid condition – see Fig. 3).

Because no effect of transparency was found, data were averaged across all opacity conditions for further analyses (see Figs. 4 and 5). Confidence intervals for the medians were computed using bootstrapping (Efron & Tibshirani, 1994) for each (eccentricity $*$ validity) condition for both OEA and VEA values.

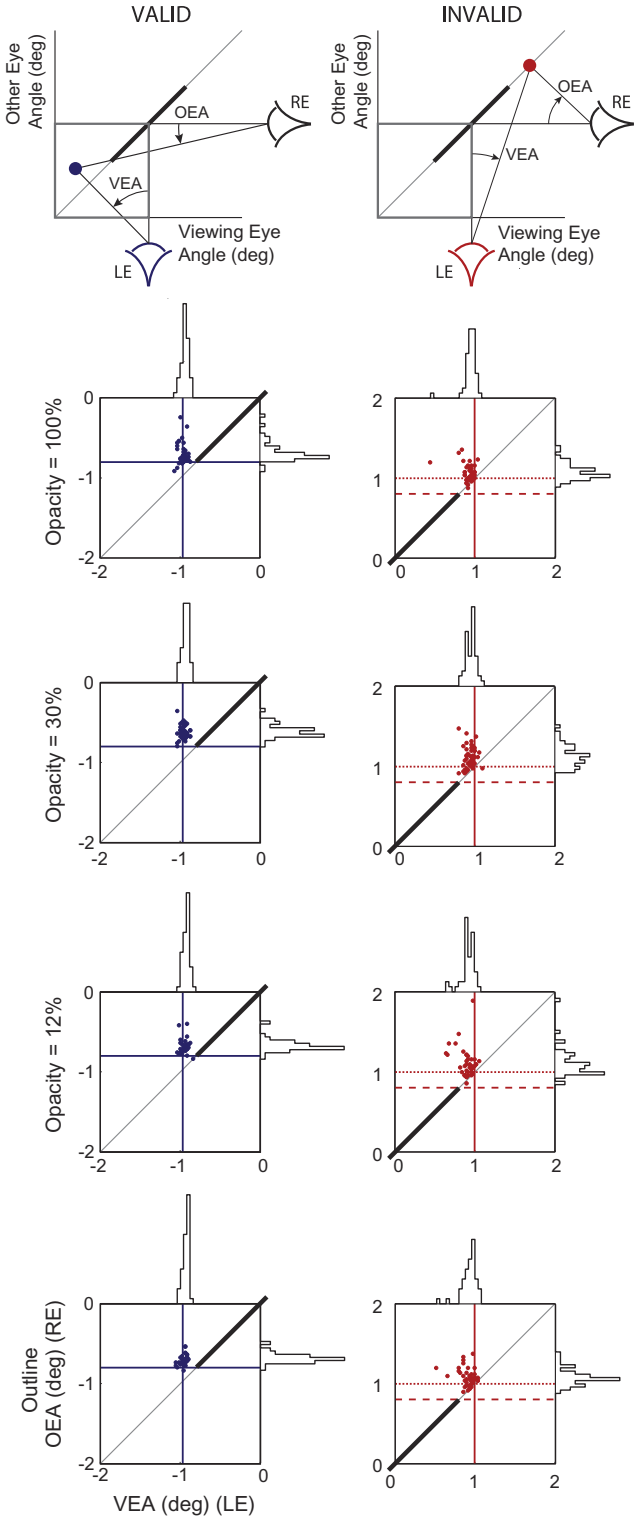


Fig. 3. Results of Experiment 1 for the opacity variable. The valid conditions (in blue) are shown in the left column and the invalid conditions (in red) in the right column. The top row illustrates the format used to plot the relationship between Viewing Eye Angle (VEA) and Other Eye Angle (OEA). The next four rows display the data for each of the four opacity conditions for the 10 arcmin eccentricity condition. Each colored dot is one percept reported by one observer. Data are pooled across all side conditions (all figures are plotted as if the monocular line were seen by the left eye). The gray diagonal line represents the zero disparity plane. The thick black line represents the position of the occluder and the colored lines show the monocular object lines of sight for both eyes and the predictions (the dotted and dashed lines represent the occlusion and double fusion predictions for the OEA respectively). The intersections of the colored lines show the different hypotheses predictions.

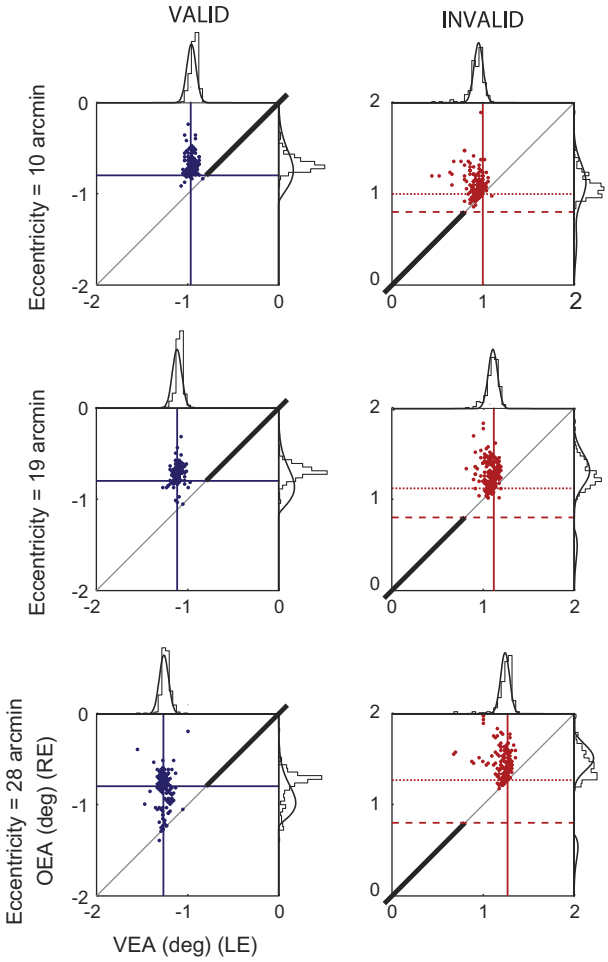


Fig. 4. Results of Experiment 1. Data are pooled across all transparency and side conditions. The three rows display the data for each of the three eccentricities of the monocular line. See legend from Fig. 3 for details.

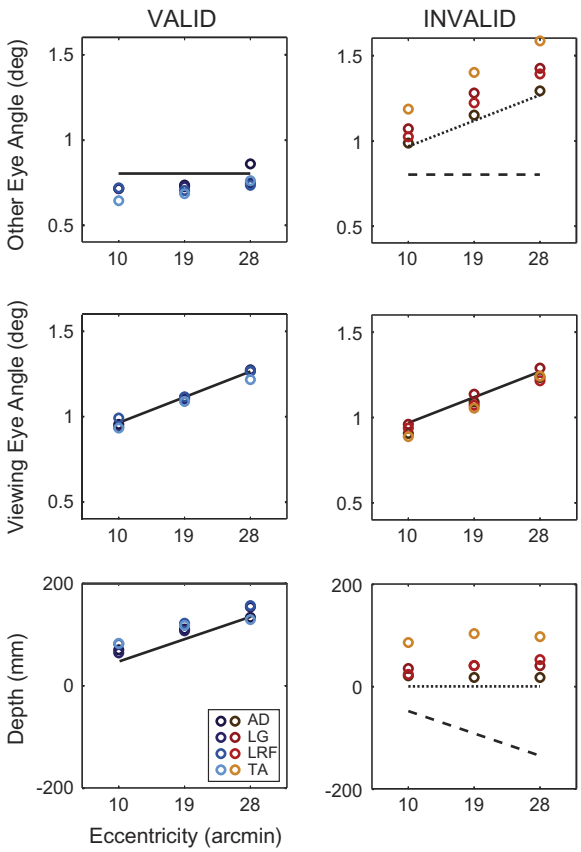


Fig. 5. Individual medians for Experiment 1. Data are averaged across all transparency and side conditions. Predictions are plotted as horizontal lines. Dashed lines represent double fusion predictions while dotted lines show occlusion predictions. Solid lines indicate same predictions for double fusion and occlusion hypotheses. Top row: Absolute values for median OEA (depth) estimations and predictions. Middle row: Absolute values for median VEA (azimuth) estimations and predictions. Bottom row: Median depth estimations in mm with and predictions. Data in blue show median estimations in the valid condition while data in red show median estimations in the invalid condition. Different shades of blue and orange represent different observers.

2.2.1.1. Valid condition. In the valid condition, the OEA values were significantly smaller than the occlusion/double fusion predictions (ω) for the three eccentricities (CI for ε_{10}: [0.684 0.723], CI for ε_{19}: [0.709 0.723], CI for ε_{28}: [0.739 0.777], prediction = 0.8). In other words, when consistent with the geometry of the scene, the line was perceived in the constraint zone. The VEA values were not different from occlusion and double fusion predictions ($\omega + \varepsilon$) for the three eccentricities (CI for ε_{10}: [0.942 0.964], prediction = 0.967; CI for ε_{19}: [1.08 1.12], prediction = 1.12; CI for ε_{28}: [1.24 1.28], prediction = 1.27), meaning that the line was perceived at the position predicted by the monocular object line of sight.

2.2.1.2. Invalid condition. The OEA values were significantly larger than the occlusion predictions ($\omega + \varepsilon$) for the three eccentricities (CI for ε_{10}: [1.033 1.080], prediction = 0.967; CI for ε_{19}: [1.22 1.27], prediction = 1.12; CI for ε_{28}: [1.38 1.44], prediction = 1.27), indicating that the monocular line was perceived behind the occluder plane. The distance between these depth estimations and the predictions tended to increase with eccentricity. The VEA values were significantly smaller than the value predicted by occlusion and double fusion ($\omega + \varepsilon$) for the 10 and 19 arcmin eccentricities (CI for ε_{10}: [0.922 0.948], prediction = 0.967; CI for ε_{19}: [1.08 1.10], prediction = 1.12) and not different from this prediction for the largest eccentricity (CI for ε_{28}: [1.23 1.27], prediction = 1.27).

2.2.2. Skewness

2.2.2.1. Other Eye Angle. In the valid condition for the 10 and 19 arcmin conditions, we observe a positive skewness (mean skewness for 10 arcmin condition = 0.726; mean skewness for 19 arcmin condition = 0.267). For the largest eccentricity we observe a negative skewness for the four observers (mean skewness for 28 arcmin condition = −0.706). In the invalid condition, the skewness of the OEA distribution is positive for all three eccentricities for the four observers (mean skewness for 10 arcmin condition = −0.894; mean skewness for 19 arcmin condition = 0.773 and mean skewness for 28 arcmin condition = 0.620).

2.2.2.2. Viewing Eye Angle. In the valid condition, the skewness of VEA distributions is very small and positive on average (mean skewness for 10 arcmin condition = 0.076; mean skewness for 19 arcmin condition = 0.152 and mean skewness for 28 arcmin condition = 0.093). The sign of this skewness means that the monocular line was perceived slightly biased toward the position of the occluder. In the invalid condition, the skewness of VEA distributions is again small but negative on average (mean skewness for 10 arcmin condition = −0.427; mean skewness for 19 arcmin condition = −0.100 and mean skewness for 28 arcmin condition = −0.401). Symmetrically, the sign of this skewness means that the monocular line was perceived slightly biased toward the position of the occluder.

2.3. Discussion of Experiment 1

2.3.1. Summary of results

The 100% opaque condition served as a classical da Vinci stereopsis baseline condition. The method of adjustment we used allowed us to collect precise estimations of the perceived line position. No effect of the opacity of the occluder was found on the perceived depth of the monocular line. For all conditions, the distribution of values for the VEA (Viewing Eye Angle, corresponding to the perceived azimuth of the monocular line) was narrowly peaked around the point predicted by the line of sight constraint but slightly asymmetric, indicating that the line was perceived slightly deviated towards the position of the occluder. On the contrary, the distribution of values for the OEA (Other Eye Angle, corresponding to the perceived depth of the monocular line) was widespread and skewed toward uncrossed disparities for the low validity conditions.

Contrary to our predictions, we found a significant effect of eccentricity in the valid condition for the OEA distribution. However, as shown in Fig. 4, this effect is small and median estimations follow predictions very closely. This effect can be attributed to a regression phenomenon previously reported by several authors (Häkkinen & Nyman, 1996; Nakayama & Shimojo, 1990). We discuss this regression in the light of a simple model in a later section.

2.3.2. No effect of transparency

All observers reported a vivid sensation of transparency and were sensitive to changes in the transmittance of the occluder. Therefore, we can assume that the opacity of the occluder was efficiently varied across the different opacity conditions.

Even though it is hazardous to assert anything from negative results, our attempts to find an effect of transparency on da Vinci stereopsis have failed. According to Nakayama and Shimojo (1990), the visual system extracts the occlusion geometry of the scene by detecting unpaired features, eye-of-origin information, depth discontinuities, object edges and opacity relationships. This geometry of occlusion is then used to determine the spatial location of these unpaired features. The experimental paradigm used by Nakayama and Shimojo (1990) did not allow them to test if da Vinci stereopsis is processed during the matching step or if the depth of the monocular object is determined once a satisfying solution to the correspondence problem has been found. These authors made no assertions about the level of processing required to compute this geometry. Our results thus suggest two alternative hypotheses. Either da Vinci stereopsis is solved before perceptual transparency is solved, or the geometry of occlusion does not include opacity information.

2.3.3. Skewness

Previous studies on da Vinci stereopsis did not dwell on the distributions of perceived depth estimations. However, the particular shape of such distributions is instructive with respect to the occlusion and double fusion hypotheses.

According to the occlusion hypothesis, an asymmetry could be expected for the OEA values in the valid condition (see Fig. 2a). In this condition, the depth estimation is constrained on one side by the minimal depth defined by the adjacent occluder's edge. In other words, this constraint forbids depth estimates that would make the line visible by both eyes, but is oblivious about depth estimates that place the line behind the occluder. The particular type of skewness we found for the OEA values in the valid condition are exactly consistent with this idea: in the 10 and 19 arcmin conditions, the distribution of OEA values had a positive skewness, extending into the occluder region. For the largest eccentricity, the mean skewness was in the other direction (negative). This spread can be explained by a phenomenon of regression to the occluder's plane (a similar interpretation was proposed by Nakayama and Shimojo (1990) and Häkkinen and Nyman (1996)). In the invalid condition, the occlusion hypothesis as stated by Nakayama and Shimojo's (1990) does not make a clear prediction with respect to the skewness of the distribution of perceived depths.

According to the double fusion hypothesis, the monocular line has a clear correspondence in the other eye (the edge of the occluder). The uncertainty in matching the monocular line with the edge should be symmetrical if matching is based on image intensity changes. However, one might argue that this uncertainty could be asymmetrical given that the monocular line can be matched with any part of the occluder. In all cases, we do not expect any change of skewness with eccentricity, or between the valid and the invalid conditions. The fact that skewness was significant in the observers' data, and that it changed across conditions, cannot be easily explained by the double fusion hypothesis.

2.3.4. Occlusion vs. double fusion

There has been an intense debate about a double fusion explanation for the phenomenon of da Vinci stereopsis (Gillam, Cook, & Blackburn, 2003; Ono et al., 1992; Pianta & Gillam, 2003). We now review how the occlusion and the double fusion hypotheses can explain our results.

Predictions following double fusion are straightforward. In both valid and invalid conditions, the perceived depth of the monocular line is computed using the distance to the occluder as disparity. If presented to the temporal side of the occluder, this disparity is uncrossed and the line is perceived further away than the occluder. Reciprocally, the line is perceived in front of the occluder when presented to the nasal side.

Predictions following the occlusion hypothesis are more complex. In the valid condition, the monocular object should be perceived behind the occluder, and therefore at a depth at least equal to the minimal depth predicted by the geometry. In the invalid condition, there is room for a symmetric interpretation where the monocular object is camouflaged by the large binocular object. However, Nakayama and Shimojo (1990) preferred the interpretation that the visual system is unable to find an adequate solution to it and thus places the monocular object at the same depth as the occluder.

Our data are more consistent with the occlusion than with the double fusion hypothesis. In the invalid condition, none of our observers perceived the monocular object in front of the occluder plane. In addition, in the valid condition, the monocular line was perceived at a depth significantly larger than the minimal depth predicted by the three eccentricities. Together with the discussion in the section above on the skewness of the distributions of perceived depths, our data therefore appear inconsistent with the double fusion hypothesis. With respect to the occlusion hypothesis, our data clearly follow the predictions in the valid condition. Indeed, the median of the perceived depth of the monocular line is behind the minimal depth imposed by the occluder, and as discussed in the section above, the interpretation of the skewness of the perceived depth distribution goes in the same direction. However, in the invalid condition, the monocular line was perceived slightly behind the occluder plane. This result is clearly inconsistent with camouflage and also deviates slightly from Nakayama and Shimojo's observations (1990). We will come back to this interpretation once we have described our simple model below.

As discussed in the introduction, different studies (Gillam, Blackburn, & Cook, 1995; Ono et al., 1992) have suggested that the depth impressions elicited by Nakayama and Shimojo' stimulus (1990) can be explained by double fusion. To address the double fusion explanation, Gillam, Cook, and Blackburn (2003) designed a da Vinci stimulus where the monocular object is a disk that cannot be "double-fused" with the adjacent edge of the occluder.

Because the results of our first experiment are neither consistent with occlusion nor with double fusion, we decided to run a second experiment to study the implication of double fusion in our stimuli.

3. Experiment 2

3.1. Methods

3.1.1. Participants

Four naïve observers (two having participated in Experiment 1) with a normal or corrected-to-normal vision were recruited in the laboratory building. All participants had experience in psychophysical observation and had normal stereo acuity and transparency sensitivity.

3.1.2. Stimulus presentation

The stereograms were presented using the same setup as for Experiment 1.

3.1.3. Stimuli

Stimuli were identical to the ones used in Experiment 1 except that the monocular line was replaced by a monocular disk (radius 0.25°) (see Fig. 6).

Experimental variables were the same as in Experiment 1. The distance between the monocular line and the black square varied randomly between three values (line eccentricity ε: 10, 19 and 28 arcmin). The eye of presentation (left or right) of the monocular line was counterbalanced and the side of presentation (left or right of the square) varied randomly to create four different conditions. The degree of opacity of the black square varied randomly between three values (100%, 30% and 12% opaque but no outline condition).

3.1.4. Procedure

As in Experiment 1, while keeping the nonius lines aligned, participants were asked to evaluate the perceived azimuth (left–right) and depth (front–back) positions of the monocular disk using an adjustment procedure. Each combination of eccentricity values, eye-of-origin, opacity values and validity configurations was repeated 12 times in total. The experiment was divided in 12 short sessions.

3.1.5. Data analysis

Data were averaged for the "side of the disk" factor to bring the total number of trials per condition to 24. As in Experiment 1, data analysis was conducted on the raw coordinates of the stereo-probe (VEA for the Viewing Eye Angle and OEA for the Other Eye Angle).

3.1.6. Predictions

If the results obtained in the first experiment are due at least partly to double fusion then we expect the depth estimations in the second experiment to be different from those the first experiment. If there is no implication of double fusion mechanisms in da Vinci stereopsis (as elicited by our stimuli), we expect the same effects as in the first experiment.

3.2. Results

As for Experiment 1, results are presented as if they resulted from the left eye condition (the disk is presented to the left eye, on the left side of the occluder in the valid condition and on the right side in the invalid condition). The distributions of OEA and VEA reports are shown in Fig. 7.

3.2.1. Main effects of experimental variables

As for Experiment 1, the normality of OEA and VEA distributions was tested using the D'Agostino normality test (D'Agostino, Belanger, & D'Agostino, 1990). Except for three OEA distributions (ε = 19 and 28 for the invalid condition and ε = 19 for the valid condition), all distributions were normal. To take into account the non-normality of a minority of OEA distributions, we conducted a repeated measures ANOVA on the median for each validity condition separately. The ANOVA conducted on the OEA measures revealed a significant effect of eccentricity ($F(2,6) = 8.34$, $P < 0.05$ for the valid condition and $F(2,6) = 0.471$, $P < 0.001$ for the invalid condition) but no effect of opacity ($F(3,9) = 2.68$, $P = 0.110$ for the valid condition and $F(3,9) = 1.733$, $P = 0.230$ for the invalid condition). The ANOVA conducted on the VEA measures revealed the same pattern of results (eccentricity: $F(2,6) = 65.7$, $P < 0.001$ for the valid condition and $F(2,6) = 69.0$, $P < 0.001$ for the invalid condition; opacity: $F(3,9) = 2.23$, $P = 0.154$ for the valid condition and $F(3,9) = 4.89$, $P = 0.028$ for the invalid condition). The ANOVA revealed a significant effect of transparency in the invalid condition. However, this effect was inconsistent across opacity conditions (the perceived horizontal position of the monocular line did not vary with a consistent pattern as opacity decreased).

Because no consistent effect of transparency was found, data were averaged across all opacity conditions for further analyses.

3.2.1.1. Valid condition. The OEA values were significantly smaller than the occlusion predictions for the 10 and 19 arcmin conditions

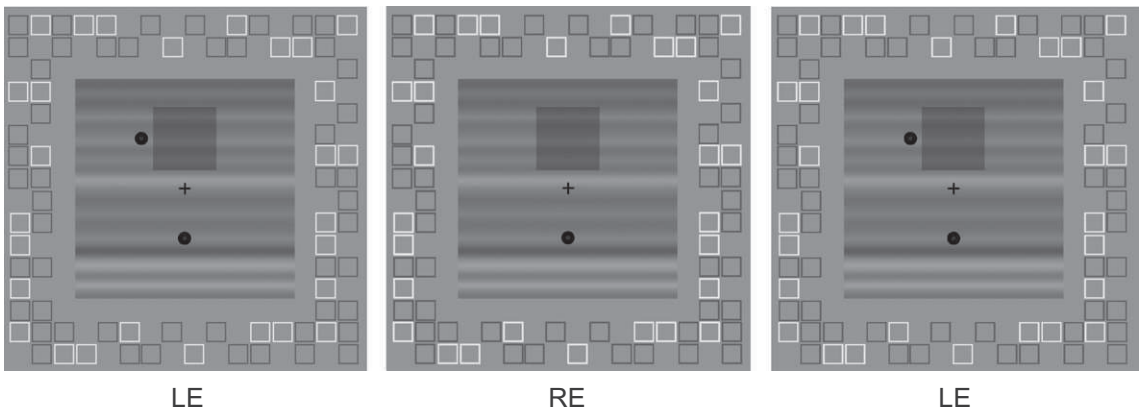


Fig. 6. Stimulus used in Experiment 2 in the 30% opaque condition (the other opacity conditions are not shown). The occlusion or valid condition can be seen by parallel-fusing the first and second columns. The monocular line is replaced by a monocular disk.

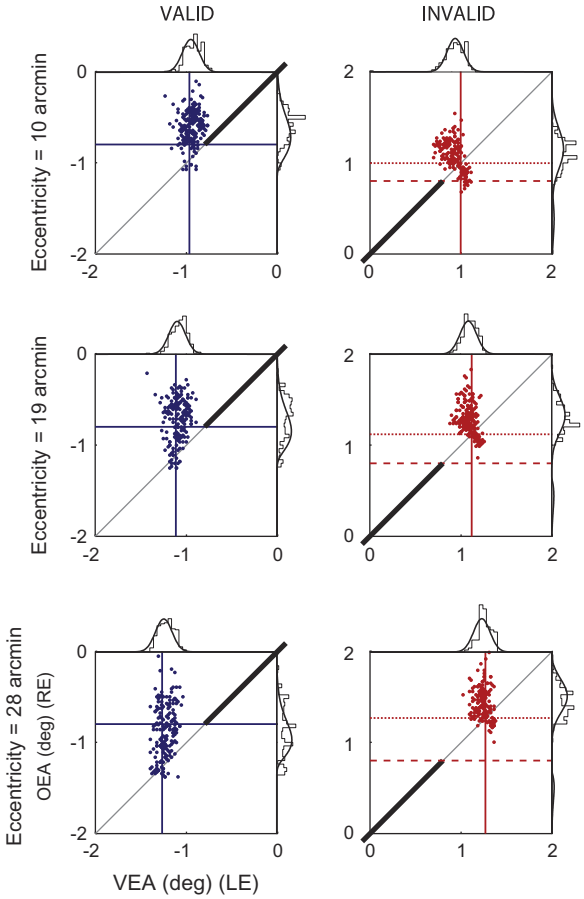


Fig. 7. Results of Experiment 2. See legend from Fig. 4 for details.

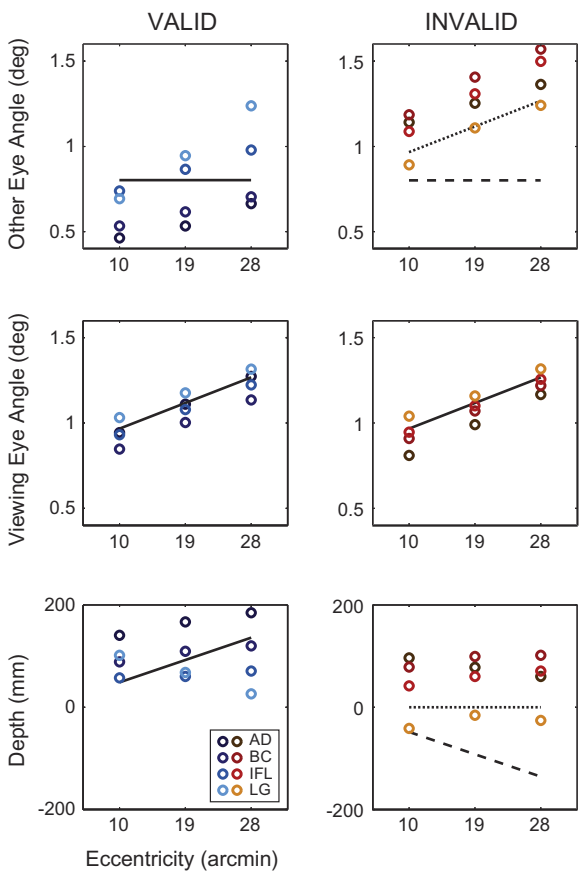


Fig. 8. Individual medians and for Experiment 2. See legend from Fig. 5 for details.

and significantly larger from the prediction for the 28 arcmin condition (CI for ε_{10}: [0.564 0.628], CI for ε_{19}: [0.631 0.717], CI for ε_{28}: [0.816 0.938]; prediction = 0.8). The VEA values were significantly smaller than both occlusion and double fusion predictions for the three eccentricities (CI for ε_{10}: [0.920 0.948], prediction = 0.967; CI for ε_{19}: [1.07 1.10], prediction = 1.12; CI for ε_{28}: [1.22 1.26], prediction = 1.27, meaning that the line was perceived closer to the occluder than the position predicted by the monocular object line of sight.

3.2.1.2. Invalid condition. The OEA values were significantly larger than the occlusion predictions for the three eccentricity values (CI for ε_{10}: [1.03 1.11], prediction = 0.967; CI for ε_{19}: [1.23 1.30], prediction = 1.12; CI for ε_{28}: [1.36 1.46], prediction = 1.27). As in the valid condition, the VEA values were significantly smaller than both occlusion and double fusion predictions for the three eccentricities (CI for ε_{10}: [0.912 0.947], prediction = 0.967; CI for ε_{19}: [1.07 1.10], prediction = 1.12; CI for ε_{28}: [1.22 1.25], prediction = 1.27.

3.2.2. Skewness

3.2.2.1. Other Eye Angle. For both valid and invalid conditions, skewness values were similar to the ones obtained in Experiment 1 but smaller: mean positive skewness for the valid condition (mean skewness = 0.386, ranging from −0.097 to 1.78) and negative skewness for the invalid condition, for all three eccentricities and the four observers (mean skewness = −0.752, ranging from −2.51 to 0.088).

3.2.2.2. Viewing Eye Angle. In the valid condition, the skewness of VEA distributions is close to zero on average (mean skewness = −0.019, ranging from −0.589 to 0.685 across observers). In the invalid condition, the skewness of VEA distributions is small but positive on average (mean skewness 0.068, ranging from −0.574 to 0.623).

3.3. Discussion of Experiment 2

The goal of Experiment 2 was to determine whether the results obtained in Experiment 1 could be (partly) explained by double fusion mechanisms. To do so, we used a monocular element (a disk) that cannot be double-fused with the edge of the occluding object.

The data obtained in this experiment were comparable to those in the first experiment, ruling out an exclusive implication of double fusion mechanisms in our stimuli. The depth and azimuth estimations in Experiment 2 are more spread than in Experiment 1 (compare Fig. 7 with Fig. 4). The greater variability for the VEA can be attributed to the fact that the disk is 2.5 times wider than the line. In contrast, the greater variability for the OEA reflects a larger proportion of estimates near the occlusion depth plane.

The most noticeable difference between the two experiments lies in the OEA measure for the valid condition (see Figs. 5 and 8, first and third rows of plots). In the first experiment, OEA measures followed the prediction patterns for the three eccentricities even though they were significantly larger. In the second experiment, OEA measures follow the prediction patterns as in the first experiment for the 10 and 19 arcmin eccentricities, but the regression observed for the 28 arcmin eccentricity is larger that in the first experiment (the perceived depth is significantly smaller than the prediction). This effect is more salient for one particular observer (shown in light blue and orange in Fig. 8). The difference between the two sets of results might

be explained by an implication of double fusion in some trials in the first experiment. Apart from these differences, both experiments provided similar results. In particular, we found a significant effect of eccentricity on OEA that corresponds to a regression to the occlusion depth plane at the largest eccentricity (28 arcmin). We now attempt to explain the effect of eccentricity on OEA with a simple model.

4. Model

Our purpose here is not to develop a complete and biologically plausible model of da Vinci stereopsis, but rather to provide a descriptive model of our results. The model includes three components that are described in more details in Appendix A. The first component characterizes the constraint provided by the edges of the occluder. This constraint favors matches inside the occluder and discourages matches outside the occluder. It is akin to a double-fusion constraint in that it allows the fusion of the monocular line with the edges of the occluder with the difference that it favors only fusion inside the object. The second component characterizes the constraint that monocular objects tend to be matched behind the object rather than in front. This constraint implements the intuitive idea of an opaque occluder that can hide any other object behind it, but precludes the possibility of camouflage. The third component is a preference for small disparities. This last component is useful to eliminate matches near the far edge of the occluder.

Overall, the combination of these three components represents the plausible locations to solve the correspondence problem when a monocular object is presented. We use the exact same model for valid and invalid conditions, the only difference being where the monocular object is presented. The model is then fitted to marginal distributions of OEA and VEA for both valid and invalid conditions, for the three eccentricities (12 distributions in total). The best fit of the model is shown as a continuous line overlaid to Figs. 3 and 6. The fitted parameters of the model are presented in Appendix A.

The model faithfully reproduces the following aspects of the data:

- in the valid condition, the distributions of OEA are skewed with a long tail extending to large depths,
- in the valid condition, we observe an increase of the spread of OEA with eccentricity,
- in the invalid condition, the distributions of OEA are closer to zero disparity than in the valid condition.

Even though the main characteristics of our data are reproduced by our model, data from Experiment 2 are better accounted for than the ones from the first experiment. For instance, the model displays more regression towards zero disparity in the first experiment than what the experimental data show. This suggests that, in the first experiment, observers may have relied on a double-fusion strategy in some trials. The stimulus in the second experiment was designed to avoid any possibility of double matching. The good match between our model and the results from our second experiment suggests that da Vinci stereopsis can be accounted for by a functional model based on scene geometry constraints, a preference for occlusion over camouflage and a prior for small disparities.

Our model implements two separate constraints for the occluder plane (a preference for occlusion over camouflage) and the fixation plane (a prior for small disparities). Although, these two depth planes were identical in our stimuli, our model makes clear predictions on the perceived position of the monocular object for a change in the occluder's depth.

5. General discussion

5.1. Summary of results from Experiments 1 and 2

We found comparable results in two experiments that used a line and a disk as monocular objects in the vicinity of an occluder. First, there was no effect of transparency on the perceived depth of the monocular object. Second, depth estimations in the valid condition were more consistent with an occlusion explanation than double fusion: the median perceived depth was within the constraint zone and the distribution of depths extended into the constraint zone (at least for small eccentricities). However, depth estimations in the invalid condition were neither in agreement with occlusion nor double fusion: the median depth was behind the occluder's plane (rather than in front) and its distribution spread over a wide range.

5.2. Implications for stereo algorithms processing unpaired features

There are two classes of strategies to infer depth for unpaired features. Monocular regions can be included at the final stages of stereo matching, to refine the disparity map (Jones & Malik, 1992): this map is processed post hoc to determine the likely localizations of depth discontinuities. In this view, occlusion relationships must be derived from the geometry of the scene before they can be integrated into the depth map. Unpaired features thus cannot be used to facilitate the construction of stereoscopic depth.

Another strategy is to postulate that there are early mechanisms capable of detecting monocular regions and occluding contours. In this view, occlusion geometry can serve as a depth cue to constrain the resolution of the matching problem (by excluding unpaired points as matching candidates) and construct the depth map of the scene. Following Nakayama and Shimojo's (1990) study, Anderson and Nakayama (1994) proposed the existence of neurons whose receptive fields are capable of sensing occlusion relationships. These occlusion relationships are extracted by hypothetical mechanisms based on eye-of-origin information and depth discontinuities. In this model, the opacity of the occluding surface is not mentioned as being critical for the processing of half-occlusion configurations. Following Anderson and Nakayama's proposal, several models postulate that the geometry of occlusion is extracted early but they differ in the mechanisms responsible for this computation (Geiger, Ladendorf, & Yuille, 1995; Grossberg & Howe, 2003; Hayashi et al., 2004; Watanabe & Fukushima, 1999). More recently, Assee and Qian (2007) pointed out the fact that these models are not parsimonious and postulate the existence of specific monocular cells. Their model is based on a simple V1–V2 feedforward structure. Depth edges and monocular regions are extracted in V2 from the outputs of V1 binocular cells.

None of the models reviewed above implement the opacity constraint as being dependent on the material properties of the occluding surface. Our results are consistent with this view and suggest that opacity, if critical for the processing of half-occlusions, is not extracted on the basis of transmittance. In this case, the opacity constraint might be achieved by implementing a simple uniqueness rule (each item from each image must be assigned at most one disparity value), as proposed by Watanabe and Fukushima (1999). This algorithm is based on the constraint that an occluding point should always exist between an unpaired point and the eye that cannot see the unpaired point.

Aside from the computational models described in this section, we propose a functional model based on the geometrical constraints of the visual scene, a bias toward occlusion rather than camouflage and a prior for small disparities. These components

can be implemented at a mid-level stage of visual processing. In this view, a general preference for small disparities is combined with the scene geometry to constrain the disparity map.

6. Conclusion

In conclusion, we failed at demonstrating that there is an interaction between perceptual transparency and da Vinci stereopsis. These results suggest that da Vinci stereopsis is solved during relatively early stages of stereoscopic processing but at the same time that it is constrained by basic geometrical information in the visual scene. By looking at the full distributions of depth and azimuth estimations rather than simply the means, we were able to describe more meticulously the percepts evoked by da Vinci stereopsis. Overall, our study questions the traditional view of stereopsis that is primarily concerned by the resolution of the correspondence problem and neglects the scene geometry.

Acknowledgments

We thank Laurie Wilcox for discussions and Michael Landy and two reviewers for their comments on an earlier draft of this manuscript. This work was supported by a Grant from the French Ministère de l'Enseignement Supérieur et de la Recherche, and by Grant ANR-2010-BLAN-1910-01 from the French Agence Nationale de la Recherche.

Appendix A

We describe here in more details the model used to determine the distributions of perceived locations of the monocular object. The model attempts to reveal all the possible locations where a monocular object could be in agreement with the occluder. In other words, we are interested in estimating the conditional probability

$$p(\mathrm{LEA}, \mathrm{REA}|\mathrm{occluder}) \tag{1}$$

where (LEA, REA) represent the coordinates (left and right eye angles) of any monocular object that can be perceived in the vicinity of the occluder. In a traditional Bayesian way, this posterior conditional distribution can be re-written as the product of a likelihood provided by the occluder and a prior expectation on the location of the monocular object (Mamassian, Landy, & Maloney, 2002)

$$p(\mathrm{LEA}, \mathrm{REA}|\mathrm{occluder}) \propto p(\mathrm{occluder}|\mathrm{LEA}, \mathrm{REA})p(\mathrm{LEA}, \mathrm{REA}) \tag{2}$$

The first term on the right-hand side of Eq. (2) represents the constraint imposed by the occluder. We assume it is the combination of two components. The first component corresponds to the constraint provided by the edges. If ω is the half-width of the occluder, then this constraint for the left eye angle (LEA) can be written

$$C_1(\mathrm{LEA}) = \left(\frac{\mathrm{LEA}-\omega}{\sigma_1^2}\right)\exp\left(-\frac{(\mathrm{LEA}-\omega)^2}{2\sigma_1^2}\right) - \left(\frac{\mathrm{LEA}+\omega}{\sigma_1^2}\right)\exp\left(-\frac{(\mathrm{LEA}+\omega)^2}{2\sigma_1^2}\right) \tag{3}$$

where σ_1^2 represents the spatial uncertainty on the edge constraint. This constraint has two parts corresponding to the left and right edges of the occluder. A similar expression applies to the right eye angle $C_1(\mathrm{REA})$.

The second component of the model favors hidden objects placed behind the occluder. It represents an opacity constraint and can be written as

$$C_2(\mathrm{LEA}, \mathrm{REA}) = -\left(\frac{\mathrm{LEA}-\mathrm{REA}}{\sigma_2^2}\right)\exp\left(-\frac{(\mathrm{LEA}-\mathrm{REA})^2}{2\sigma_2^2}\right) \tag{4}$$

where σ_2^2 represents the spatial uncertainty on the opacity constraint. The edge and opacity constraints combine to provide an overall constraint provided by the occluder. We take this combination to be a weighted sum where a weight α is assigned to the opacity constraint. The overall constraint provided by the occluder is therefore

$$p(\mathrm{occluder}|\mathrm{LEA}, \mathrm{REA}) \propto \lfloor C_1(\mathrm{LEA}) + C_1(\mathrm{REA}) + \alpha C_2(\mathrm{LEA}, \mathrm{REA}) \rfloor \tag{5}$$

where the symbols $\lfloor\rfloor$ indicate that we take only the positive part of this combination.

The third component of the model is a prior for small disparities

$$p(\mathrm{LEA}, \mathrm{REA}) \propto \exp\left(-\frac{(\mathrm{LEA}-\mathrm{REA})^2}{2\sigma_3^2}\right) \tag{6}$$

where σ_3^2 characterizes the strength of the zero disparity constraint. This prior constraint is combined with the overall occluder constraint (Eq. (5)) according to Eq. (2). The proportional sign in that equation corresponds to the fact that the product has to be normalized so that the posterior is a probability distribution (*i.e.* sums to 1; see Mamassian, Landy, & Maloney, 2002).

All together, the occluder constraint and the prior for small disparities define the locations in binocular space where a monocular object can be seen in the vicinity of the occluder. We have represented these locations in Fig. 9, where for the purpose of the illustration, we have preserved the negative parts of the occluder computation in Eq. (5). We note that the areas where a monocular line can easily be matched (in orange) are behind the occluder, as well as slightly to the left of the occluder for the left eye and slightly to the right for the right eye. In contrast, there are two inhibitory zones (in blue) on either side of the occluder. These inhibitory zones are responsible for the skewness of the distribution of reported depth of the monocular objects in our data.

To obtain quantitative predictions for the monocular line or disk stimuli, we assume that these stimuli are located with their own uncertainty

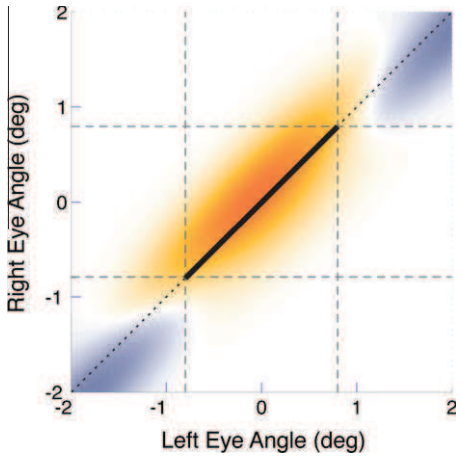


Fig. 9. Modeled constrained space by the occluder. The occluder is shown as a thick black diagonal line between −0.8° and +0.8° in both eyes, thus perceived as a fronto-parallel rectangle of width 1.6°. The model attempts to reveal the locations in binocular space where an object presented monocularly could be perceived in agreement with the occluder. Orange locations indicate positive areas, namely locations where a monocular object could indeed be matched. Blue locations indicate negative areas, namely locations where correspondence would be inhibited. See Appendix A for model details.

Table 1
Parameters of the model adjusted to the experimental results.

	σ_1 (deg)	σ_2 (deg)	α	σ_3 (deg)	σ_4 (deg)
Experiment 1 (line)	1.06	0.35	0.42	0.33	0.051
Experiment 2 (disk)	1.12	0.24	0.49	0.36	0.093

$$M(\mathrm{VEA}) = \exp\left(-\frac{(\mathrm{VEA} - (\omega + \varepsilon))^2}{2\sigma_4^2}\right) \quad (7)$$

where $(\omega + \varepsilon)$ is the physical location of the monocular object (when it is left of the occluder) and σ_4^2 characterizes its spatial uncertainty. This latter parameter can be adjusted to take into account the width of the monocular object (a wider object – e.g. a disk compared to a line – carries more spatial uncertainty). This monocular object constraint is combined with the posterior distribution by taking their product. In the end, we obtain as a model

$$p(\mathrm{VEA}, \mathrm{OEA}) \propto p(\mathrm{VEA}, \mathrm{OEA}|\mathrm{occluder})M(\mathrm{VEA}) \quad (8)$$

where the proportional sign is again used here to guarantee a probability distribution function for possible pairs of VEA and OEA associated to a specific monocular object.

The exact same model is used for valid and invalid conditions, the only difference being the location of the monocular object. From the model, we extract the distributions of VEA and OEA for each of the six experimental conditions (valid and invalid locations of the monocular object for the three eccentricities). We then adjust the five parameters of the model to minimize the squared distance between the predicted distributions and the data. The fitted parameters of the model are presented in Table 1 and the best fitted distributions are superimposed onto Figs. 4 and 7.

References

Anderson, B. L. (2008). Transparency and occlusion. In A. I. Basbaum, A. Kaneko, G. M. Shepherd, & G. Westheimer (Eds.) & T. D. Albright & R. Masland (Vol. Eds.), *The senses: A comprehensive reference* (Vol. 2, Vision II, pp. 239–244). San Diego: Academic Press.

Anderson, B., & Nakayama, K. (1994). Toward a general theory of stereopsis: Binocular matching, occluding contours, and fusion. *Psychological Review, 101*(3), 414–445.

Assee, A., & Qian, N. (2007). Solving da Vinci stereopsis with depth-edge-selective V2 cells. *Vision Research, 47*(20), 2585–2602.

Backus, B. T., Fleet, D. J., Parker, A. J., & Heeger, D. J. (2001). Human cortical activity correlates with stereoscopic depth perception. *Journal of Neurophysiology, 86*, 2054–2068.

Brainard, D. (1997). The psychophysics toolbox. *Spatial Vision, 10*(4), 433–436.

Cook, M., & Gillam, B. (2004). Depth of monocular elements in a binocular scene: The conditions for da Vinci stereopsis. *Journal of Experimental Psychology: Human Perception and Performance, 30*(1), 92–103.

D'Agostino, R. B., Belanger, A., & D'Agostino, R. B. J. (1990). A suggestion for using powerful and informative tests of normality. *The American Statistician, 44*(4), 31–321.

Efron, B., & Tibshirani, R. J. (1994). *An introduction to the bootstrap*. New York: Chapman & Hall.

Geiger, D., Ladendorf, B., & Yuille, A. (1995). Occlusions and binocular stereo. *International Journal of Computer Vision, 14*, 211–226.

Gillam, B., Blackburn, S., & Cook, M. (1995). Panum's limiting case: Double fusion, convergence error, or 'da Vinci stereopsis'. *Perception, 24*(3), 333–346.

Gillam, B., & Borsting, E. (1988). The role of monocular regions in stereoscopic displays. *Perception, 17*, 603–608.

Gillam, B., Cook, M., & Blackburn, S. (2003). Monocular discs in the occlusion zones of binocular surfaces do not have quantitative depth – A comparison with Panum's limiting case. *Perception, 32*(8), 1009–1019.

Grossberg, S., & Howe, P. (2003). A laminar cortical model of stereopsis and three-dimensional surface perception. *Vision Research, 43*(7), 801–829.

Häkkinen, J., & Nyman, G. (1996). Depth asymmetry in da Vinci stereopsis. *Vision Research, 36*(23), 3815–3819.

Harris, J. M., & Wilcox, L. M. (2009). The role of monocularly visible regions in depth and surface perception. *Vision Research, 49*, 2666–2685.

Hayashi, R., Maeda, T., Shimojo, S., & Tachi, S. (2004). An integrative model of binocular vision: A stereo model utilizing interocularly unpaired points produces both depth and binocular rivalry. *Vision Research, 44*(20), 2367–2380.

Howard, I. P., & Rogers, B. J. (2002). *Seeing in depth* (Vol. II). Porteous Press.

Hubel, D. H., & Wiesel, T. N. (1959). Receptive fields of single neurons in the cat's striate cortex. *Journal of Physiology, 148*, 574–591.

Jones, J., & Malik, J. (1992). *A computational framework for determining stereo correspondence from a set of linear spatial filters*. Berkeley: University of California, Computer Science Division (EECS).

Lawson, R., & Gulick, W. (1967). Stereopsis and anomalous contour. *Vision Research, 7*(3), 271–297.

Mamassian, P., Landy, M. S., & Maloney, L. T. (2002). Bayesian modelling of visual perception. In R. Rao, B. Olshausen, & M. Lewicki (Eds.), *Probabilistic models of the brain: Perception and neural function* (pp. 13–36). Cambridge, MA: MIT Press.

Metelli, F. (1985). Stimulation and perception of transparency. *Psychological Research, 47*(4), 185–202.

Nakayama, K., & Shimojo, S. (1990). Da Vinci stereopsis: Depth and subjective occluding contours from unpaired image points. *Vision Research, 30*(11), 1811–1825.

Ono, H., Shimono, K., & Shibuta, K. (1992). Occlusion as a depth cue in the Wheatstone–Panum limiting case. *Perception and Psychophysics, 51*(1), 3–13.

Panum, P. L. (1858). Untersuchungen uber das Sehen mit Zwei Augen. *Kiel*.

Parker, A. J. (2007). Binocular depth perception and the cerebral cortex. *Nature Review Neuroscience, 8*(5), 379–391.

Pelli, D. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*(4), 437–442.

Pianta, M., & Gillam, B. (2003). Paired and unpaired features can be equally effective in human depth perception. *Vision Research, 43*(1), 1–6.

Porter, T., & Duff, T. (1984). Compositing digital images. *Computer Graphics, 18*(3), 253–259.

Preston, T. P., Li, S., Kourtzi, Z., & Welchman, A. E. (2008). FMRI selectivity for perceptually-relevant binocular disparities in the human brain. *Journal of Neuroscience, 28*, 11315–11327.

Qiu, F., & Von Der Heydt, R. (2007). Neural representation of transparent overlay. *Nature Neuroscience, 10*(3), 283–284.

Sekuler, A. B., & Palmer, S. E. (1992). Perception of partly occluded objects: A microgenetic analysis. *Journal of Experimental Psychology: General, 121*(1), 95–111.

Singh, M., & Anderson, B. (2002). Toward a perceptual theory of transparency. *Psychological Review, 109*(3), 492–519.

Watanabe, O., & Fukushima, K. (1999). Stereo algorithm that extracts a depth cue from interocularly unpaired points. *Neural networks: The official journal of the International Neural Network Society, 12*(4–5), 569–578.

V Using sound as a tool to study motion-in-depth

1 Introduction

In order to hit a tennis ball with a racket, the player must calculate its trajectory and direction with accuracy. Motion-in-depth can be extracted from multiple cues such as optic flow, retinal image expansion or binocular disparity. When an object moves in depth, its disparity relative to the fixation plane changes over time. For an approaching or a receding object, its image will move in opposite or equal directions on the two eyes' retinae. The term stereomotion refers to the perception of motion-in-depth defined exclusively by binocular information.

Visual scientists started to get interested in motion-in-depth in the early seventies. At this time, there was a general trend to define the global functioning of the visual system: visual attributes are first segregated and processed in highly specified cortical areas and finally reintegrated together to form a coherent interpretation of the visual scene. From this point of view, motion-in-depth appeared to be a challenging case study. While it requires the integration of motion and disparity information together, psychophysical (Regan & Beverley, 1973a) and neurophysiological (Regan & Beverley, 1973b) evidence suggested very early on that motion-in-depth constitutes a full independent visual attribute (Tyler, 1971).

In the present section, we will review the current literature on the field, focusing on methodological aspects of particular interest for the experimental work presented in sections 2 & 3 of this chapter. In addition, we will highlight issues and questions that are still to be addressed.

1.1 Two independent cues for stereomotion

In this section, we will discuss the existence of two independent cues for the perception of motion-in-depth, their respective sensitivities and their relative utilities.

1.1.1 Definitions

Rashbass & Westheimer (1961) were the first to postulate the existence of two independent cues to track the position in depth of objects (Fig V.1). According to them, this could be achieved either by recording "the rate of change of the difference in the position of the images in the two eyes" or by computing "the difference between the velocity of the movement of the two images across the two retinae". These two sources of information are now referred to as *change of disparity over time* (CDOT) and *interocular velocity difference* (IOVD).

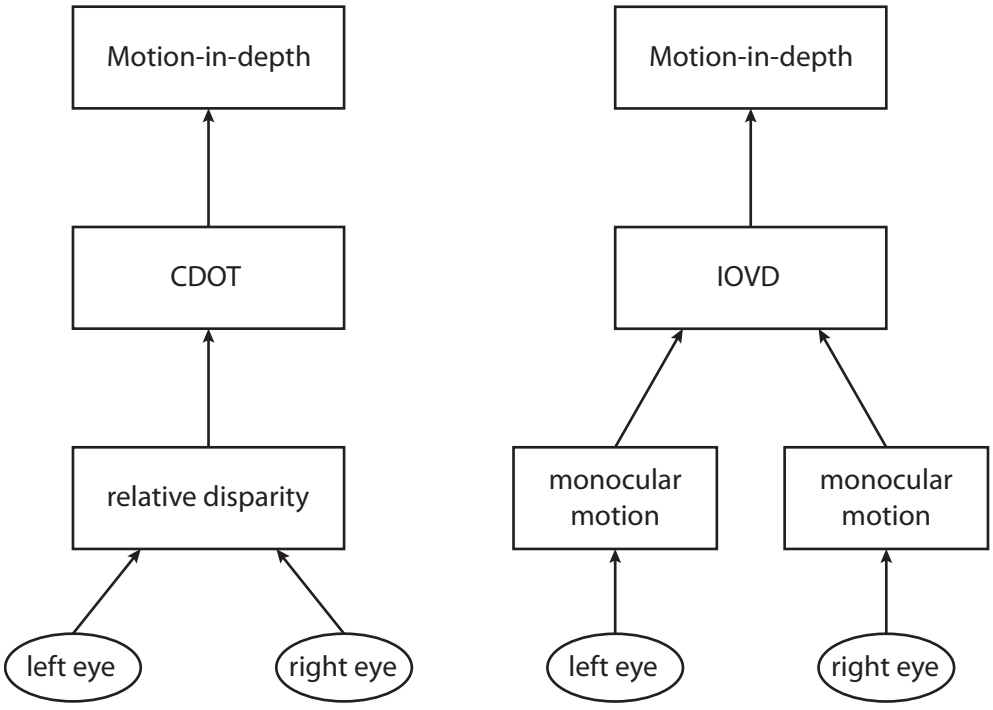


Figure V.1 | Illustration of how the two cues to motion-in-depth are thought to be processed.

CDOT

One possibility for the visual system is to extract the binocular disparity of an object relative to the fixation plane and to track how this information varies over time. Speed information can be extracted from the amount of disparity change over time while the difference between lateral displacements between the two eyes' images informs on the direction of motion.

IOVD

Another possibility is to rely on the velocity of the image of an object extracted separately for the two eyes' images. The speed and direction of motion-in-depth can then be computed by taking the ratio between the velocities in the two monocular motion components.

In ecological viewing situations, CDOT and IOVD are always present and vary congruently. Since the early nineties, research on stereomotion has focused on designing new stimuli to isolate the two sources of information in order to characterize their processing and utilities. Relying on existing knowledge and methodology developed for the study of static stereopsis, early studies have focused on understanding the role of CDOT, sometimes underestimating the importance of IOVD. However, considerable progress has been made during the past decade on understanding the role of IOVD using complex motion stimuli (Brooks, 2002a; Rokers, Czuba, Cormack, & Huk, 2011). Recently, the development of imaging techniques has allowed researchers to better understand the mechanisms underlying the processing of CDOT and IOVD (Likova & Tyler, 2007; Rokers et al., 2009).

1.1.2 Understanding the role of CDOT and IOVD for *detection* of motion-in-depth

Early psychophysical work on motion-in-depth has focused on determining the conditions necessary for the *detection* of motion-in-depth and found that the presence of CDOT was critical for the perception of 3D motion. However, more recent studies have convincingly demonstrated that IOVD alone was sufficient for the detection of motion-in-depth.

1.1.2.1 Isolating the change of disparity over time cue

To isolate the binocular components of motion-in depth, Regan (Regan, 1993) designed an original stimulus called the dynamic random dot stereogram (DRDS — Fig. V.2), based on the random dot stereogram (RDS — Fig. V.2) developed by Julesz (1964b), see chapter III, section 2.1.2). In a DRDS, a new random dot pattern is generated on each new video frame. Stereopsis is obtained as in classic RDS by adding an offset between the two eyes' images. To obtain motion-in-depth, this offset is systematically increased or decreased on each new video frame. When a CDOT is applied on a classic static RDS, a portion of the image can clearly be seen moving laterally on each monocular image. In DRDS, the entire dot pattern is refreshed every frame, creating random correlations across frames resulting in the perception of motion in all directions at random speeds. Yet, the visual system is still able to detect the systematic lateral displacement applied to the portion of interest and to use it to compute its disparity and track the changes of disparity over time. In other words, the use of DRDS preserves the CDOT information while making the IOVD cue inconsistent and thus unusable. Similarly to Beverley & Regan (Beverley & Regan, 1973), Regan (1993) manipulated the ratio between the amount of lateral displacement in the two eyes' images. This resulted in an apparent change in the perceived direction of motion-in-depth, demonstrating that the CODT is sufficient to detect motion-in-depth.

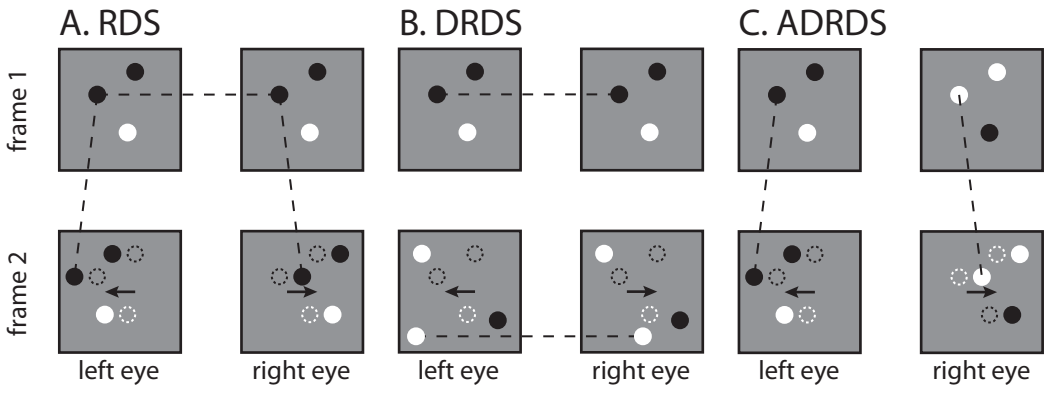


Figure V.2 | Schematic illustration of the stimuli used to isolate CDOT and IOVD. A. Random dot stereogram display containing both temporal and spatial correlations (CDOT and IOVD). B. Dynamic random dot stereogram. A new pattern of dots is

generated every frame producing random motion signals. A disparity can be applied to this new pattern, CDOT information can thus be conveyed without a clear monocular motion pattern. C. Anticorrelated random dot stereogram. Each dot in one eye has a reversed polarity in the other eye. ARDSs elicit no percept of depth but convey monocular motion signals (IOVD).

Cumming & Parker (1994) measured thresholds for the detection of disparity modulations for both RDS and DRDS stimuli. They found that these thresholds were equally low for both types of stimuli, suggesting that CDOT is sufficient to detect motion-in-depth. Furthermore, the authors argued that there was no experimental evidence of the implication of IOVD.

1.1.2.2 Isolating the interocular velocity difference cue

Clinical evidence

Two clinical studies have reported the existence of motion-in-depth without static stereopsis in strabismic patients, suggesting the co-existence of independent CDOT and IOVD information. First, Kitaoji and Toyama (1987) showed selective preservation of motion-in-depth or static stereopsis for strabismic patients. Later, in a similar study, Maeda, Sato, Ohmura, Miyazaki, Wang & Awaya (1999) showed that more than half of their patients who did not have stereopsis reported seeing motion-in-depth.

Motion aftereffects

The studies reported above used stimuli containing both CDOT and IOVD to show that IOVD alone could elicit motion-in-depth for patients who were not sensitive to CDOT and that monocular velocity adaptation could produce motion-in-depth. To extend these findings to a healthy population, motion aftereffects have proven to be an efficient tool. Monocular motion adaptation has been shown to affect motion-in-depth perception, implying the existence of a velocity-based cue.

To investigate the relative contributions of CDOT and IOVD, Brooks (Brooks, 2002b) used a cue conflict paradigm. In two separate experiments he manipulated independently the direction information given by CDOT and IOVD. In a first experiment, he manipulated the 3D trajectory information carried by IOVD by producing a velocity aftereffect in one eye. In a second experiment, the author took

advantage of the fact that the perceived direction of a binocularly defined stimulus is systematically biased by ocular dominance. This bias was used to differentiate the 3D direction information computed from CDOT and IOVD. Brooks found that the perceived trajectory of 3D motion was affected by the velocity aftereffect in Experiment 1 but not by the interocular dominance in Experiment 2, suggesting that 3D trajectory is extracted from the IOVD cue.

Similarly, Fernandez & Farrell (2005) showed that adapting to a frontoparallel motion (seen binocularly) improved motion-in-depth direction discrimination compared to adapting to random noise or to a static display *only* when the stimulus contained IOVD information. When the stimulus contained only CDOT, speed sensitivity was worse. This opposite effect suggested a significant contribution of IOVD to the perception of motion in depth.

However, Shioiri, Kahehi, Tashiro & Yaguchi (2009) pointed out that the test stimuli used by Brooks (2002b) and Fernandez & Farrell (2005) contained disparity cues and argued that motion aftereffects could have influenced the perception of motion-in-depth through disparity processing. To circumvent this issue, the authors successfully measured the occurrence of a perception of motion-in-depth in a static display after a lateral motion adaptation period, confirming that IOVD alone can support motion-in-depth.

Cancelation of CDOT

In order to isolate the IOVD information, all disparity information must be removed while the correlation of the dots' positions over time is preserved.

To achieve this dichotomy, Shioiri, Saisho & Yaguchi (2000) used binocularly uncorrelated random-dot kinematograms. The kinematograms contained two frames for each eye. The left and right images were uncorrelated, providing no binocular information for disparity processing. Each image was displaced in opposite directions in the two eyes between the first and second frame, providing motion-in-depth signals. By presenting only two frames, the authors sought to minimize the possibility of spurious correlations between the left and right images, which could have been used to extract disparity (hence CDOT). Results showed that the observers' ability to judge the relative direction of motion in the kinematogram was above chance. To rule out any remaining possible effect of random binocular pairing, the authors spatially

separated the right and left eye images in adjacent horizontal bands. Again, they showed that the direction of motion could be identified, even without any binocular overlap. It should be noted that Allison, Howard & Howard (1998) and Harris, Nefs & Grafton (2008) stated that separating the two eyes' information horizontally could still produce spurious disparities at bands' boundaries.

To circumvent the issue of spurious disparities in uncorrelated displays, Rokers, Cormack & Huk (2008) employed dynamic anticorrelated random dot stereograms (ARDS, Fig. V.2). This type of displays have been shown to produce no perception of depth (Cumming & Parker, 1997) even though they produce clear activation of disparity sensitive neurons in the area V1 of macaque monkeys. Rokers and colleagues varied the degree of contrast correlation between the two eye's images and showed that when the RDSs were anticorrelated, static depth perception was substantially impaired while motion-in-depth was unimpaired through all polarity-correlation conditions. Their results strongly support the idea that IOVD alone is sufficient for the perception of motion-in-depth. Furthermore, their data suggests that the disparity information required to track motion-in-depth cannot be derived from the raw activity of V1 disparity sensitive cells.

Recently, the same research group (Rokers et al., 2011) conducted a series of experiments to determine the nature of the motion information implicated in the computation of IOVD and the level of processing required for IOVD based motion-in-depth. These authors used motion stimuli called "plaids" in which two superimposed sinusoidal gratings drifting in different directions (or "component directions") produce a plaid pattern that is perceived as moving in a single coherent direction (or "pattern motion direction"). It has previously been shown that while component motion signals are processed in V1, "pattern motion" neurons found in MT are sensitive to the direction of the pattern motion, regardless of the direction of the component motions. The authors found that motion-in-depth sensitivity depended on the exclusively pattern motion and not the component motions.

1.1.3 Understanding the role of CDOT and IOVD for *discrimination* of speed of motion-in-depth

A majority of the studies aiming at characterizing the relative importance and utility of CDOT and IOVD, focused on the conditions necessary for the *detection* of motion-in-depth. A parallel line of work aimed at understanding the role of CDOT and IOVD by looking at speed *discrimination*. To anticipate, these studies have found that IOVD plays a major role in speed discrimination.

In 1995, Harris & Watamaniuk (H1995) conducted a series of experiments to investigate whether there existed a system exclusively dedicated to processing the speed of motion-in-depth, and if so, whether it required the use of CDOT or IOVD or both. In a first experiment, they measured Weber fractions for discriminating the speed in 3D motion stimuli containing both CDOT and IOVD and 2D motion stimuli consisting of the right eye image of the 3D motion stimuli. They found that the Weber fraction was comparable in the two motion conditions. In a second experiment, the authors used a DRDS to isolate the CDOT component and found that Weber fractions were at least twice as large as in the 3D motion condition (containing both CDOT and IOVD) from the first experiment, suggesting that CDOT is not useful for computing the speed of motion-in-depth. However, when they examined performance for these DRDS as a function of the stimulus duration, they found that long stimuli were perceived faster than shorter stimuli, suggesting that a comparison between static disparities at the beginning and end of trials was used to extrapolate speed. The authors concluded that the observers might have not based their judgments on the actual speed.

Harris & Watamaniuk's (1995) DRDS stimuli moved from away from the observer, passing through the plane of zero disparity and thus becoming momentarily invisible to the stereo system. Portfors-Yeomans and Regan (1996) claimed that this difference in detectability is critical to interpret Harris & Watamaniuk's results. To test this possibility, these authors ran a similar experiment to Harris & Watamaniuk's at different disparity pedestals and found similar Weber fractions for cyclopean stimuli (DRDS, CDOT only) and monocularly visible stimuli (CDOT + IOVD). However, as pointed out by Brooks & Stone (2004), it is not clear whether this

difference in performance for CDOT-only stimuli between the two studies is due to the visibility difference or to the addition of a pedestal.

Brooks (2002a) used the velocity aftereffect to investigate 3D speed perception and address issues raised by methodological aspects of both Harris & Watamaniuk (1995) and Portfors-Yeomans & Regan (1996) studies. In a series of experiments, the author induced a velocity aftereffect by adapting observers to either classic RDS (containing both CDOT and IOVD) or to uncorrelated RDS (containing only IOVD). First, he showed that adaptation to classic and uncorrelated RDS produced a velocity aftereffect of identical strength when the motion passed through the plane of zero disparity, strongly suggesting that the CDOT component is not used to compute the speed of motion in depth for motion located around the fixation plane. In contrast, he showed that when motion-in-depth did not cross this area, classic RDS containing CDOT and IOVD produced a stronger aftereffect than uncorrelated RDS, suggesting that CDOT had a substantial influence on speed computation. By showing that CDOT and IOVD are used differently to compute speed depending on whether the motion passes through the fixation plane, Brooks reconciled the apparently conflicting results of Harris & Watamaniuk (1995) and Portfors-Yeomans and Regan (1996).

To examine in more details the effect of a disparity pedestal on 3D speed processing, Brooks & Stone (2004) measured speed discrimination thresholds at different disparity pedestals for both RDS and DRDS stimuli. He found no effect of the disparity pedestal and that thresholds for DRDS were on average 1.7 times higher than for RDS, even though stereoacuity was equally good for these two types of stimuli. These results suggest that even though CDOT can be used to compute speed, IOVD provides a more precise cue to motion-in-depth speed perception.

Brooks (2001) also used luminance contrast to address the issue of speed computation and found that the "Thompson effect" (a reduction in contrast leading to a reduction in perceived speed) was present in similar proportions in both 2D and 3D motion perception. In line with other work from this author, this result suggests that monocular motion is the dominant input to speed computation in motion-in-depth.

1.2 Utility of CDOT and IOVD information

Since CDOT and IOVD are both present natural scenes, it is reasonable to hypothesize that these two types of information might have different and complementary utilities depending on the stimuli and/or the task (see Harris, Nefs & Grafton (2008) for a more exhaustive review).

1.2.1 Motion detection versus Speed discrimination

Several studies, for example Cumming (1995) showed that thresholds for the detection of motion-in-depth correlated with static stereoacuity and became worse as the disparity pedestal increased. On the contrary, Brooks (2002a; 2004) reported that speed discrimination thresholds were worse for CDOT-only stimuli and that they did not depend on the pedestal, suggesting that speed discrimination might rely on a mechanism that is insensitive to the disparity pedestal.

1.2.2 Relative use of CDOT and IOVD across the visual field

In a study detailed in the above section, Kitaoji & Toyama (1987) tested strabismic patients and found that the preservation of central and peripheral motion-in-depth and static stereopsis could occur independently. In the same line of work, Czuba, Rokers, Huk, & Cormack (2010) measured direction-discrimination sensitivity different types of motion-in-depth stimuli. They found that close to the fovea and for the slowest speeds, sensitivity was highest for the CDOT-only stimuli and lowest for the IOVD-only stimuli. Increasing eccentricity reversed the sensitivity pattern for both types of stimuli and increasing speed clearly reversed the sensitivity pattern for the CDOT-only stimuli and had a mixed effect for IOVD-only stimuli. The CDOT + IOVD sensitivity pattern was identical to the IOVD-only one, strongly implying that outside the fovea, the visual system relies primarily on IOVD cues to compute motion-in-depth.

A study by Brooks & Stone (2004) examined the spatial scale of the mechanisms supporting the computation of CDOT and IOVD and found that the spatial resolution of the CDOT mechanism (and the static disparity system) was on average

nine times coarser than the IOVD mechanism (and the monocular motion system). This finding gives strong evidence for the benefit of having two independent sources of information for computing motion-in-depth.

1.3 Evidence for specific motion-in-depth mechanisms

Research studies described in the above section have clearly established that motion-in-depth can be extracted from two independent sources of information, namely CDOT and IOVD. Another stream of research has focused on understanding whether these signals are combined together by 3D-motion specialized mechanisms or if CDOT and IOVD are processed independently by static disparity detectors and by 2D motion sensitive neurons.

1.3.1 Evidence from sensitivity measures

To address this question, Tyler (1971) used a stimulus consisting of two lines moving either in identical or in opposite directions and showed that sensitivity to stereoscopic motion-in-depth (i.e. when the two lines moved in opposite directions) was reduced compared to monocular lateral motion (i.e. when the two lines moved in identical directions). This sensitivity discrepancy can be considered as the first evidence of the existence of distinct mechanisms for the computation of 2D and 3D motion.

1.3.2 Evidence from adaptation studies

In 1973, Beverley & Regan (1973) demonstrated the existence of specific mechanisms for the processing of motion-in depth by showing that adaptation to motion-in-depth was independent of adaptation to static disparities. More specifically, they showed that adaptation was selective to the direction of motion, suggesting the existence of neural mechanisms sensitive selectively to the direction of motion-in-depth and not only to the monocular components of their stimuli.

Shioiri, Kahehi, Tashiro & Yaguchi (2009) compared the spatial frequency dependence between 2D and 3D motion aftereffects to assess the level of processing

required for motion-in-depth. It has been shown that the motion aftereffect is optimal when the spatial frequencies of the adaptation and test stimuli are identical. The authors found that the 3D motion aftereffect did not much depend on spatial frequency, implying the existence of a motion integration step previous to the calculation of interocular velocity differences. This difference in processing between lateral motion and motion-in-depth suggests that 2D and 3D motion are processed independently.

Czuba, Rokers, Guillet, Huk & Cormack (2011) first compared the effect of adaptation of 2D or 3D motion and found that large 3D motion aftereffects that could not be explained by a simple combination of monocular aftereffects. This result allowed them to confirm the existence of neurons specifically tuned to 3D motion. In a second experiment, they measured 3D motion aftereffects of stimuli containing exclusively CDOT or IOVD and found a small aftereffect in the CDOT condition while the aftereffect in the IOVD condition was as large as the aftereffect reported in the first experiment. This difference confirmed the central role of IOVD in motion-in-depth processing. The results of Czuba et al. are in line with those reported by Brooks (2002a) who found a larger velocity aftereffect for motion-in-depth than for monocular lateral motion.

1.3.3 Evidence from other psychophysical studies

Harris & Watamaniuk (1995) and Brooks & Stone (2004) tested whether speed was computed by judging the velocity of only one monocular motion signal or by combining monocular motion signals from the two eyes and showed that a comparison of monocular motion cues was used to discriminate the speed of motion-in-depth.

To investigate how monocular motion signals are combined to produce motion in depth, Rokers Czuba Cormack & Huk (2011) designed stimuli in which motion signals were carried by small Gabors that could not be matched binocularly due to large spatial separations within and between the two eyes. Yet, these stimuli elicited a clear percept of motion-in-depth, implying that the eye-of-origin information can be

recovered by non-conventional stereo mechanisms and incorporated in later motion processing to compute motion-in-depth.

1.3.4 Evidence from neurophysiology recordings

Zeki (1974) and Poggio & Talbot (1981) provided early evidence of the existence of neurons tuned for motion-in-depth in area MT of the cortex of the Rhesus Monkey. Similarly, Cynader & Regan (1978) record motion-in-depth sensitive neurons in the area 18 of the Cat's cortex. However, Maunsell and van Essen (1983) found no evidence of true motion-in-depth sensitivity in MT and stated that previous findings might have confounded motion-in-depth and mere disparity sensitivity. Later, Cynader & Regan (1982) and Spileers, Orban, Gulyas & Maes (1990) reported neurons on the area 18 of the Cat's cortex having motion-in-depth sensitivity that did not change with disparity.

1.3.5 Evidence from imaging studies

Neurophysiological evidence collected on the Cat and the Monkey supports the idea that stereomotion is processed together with lateral motion and disparity in the area 18 of the Cat's cortex and in area MT of the Monkey's cortex. Likova & Tyler (2007) used functional magnetic resonance imaging to locate the processing of stereomotion in the human brain. They used DRDS stimuli to isolate the CDOT component of motion-in-depth and found that these cyclopean stimuli generated specific activation in a region anterior to the hMT+ complex (human homolog of the Monkey's MT — Fig. V.3). This finding supports the idea that stereomotion processing takes place in a specialized cortical area, adjacent to hMT+ and is therefore complementary but distinct from lateral motion processing.

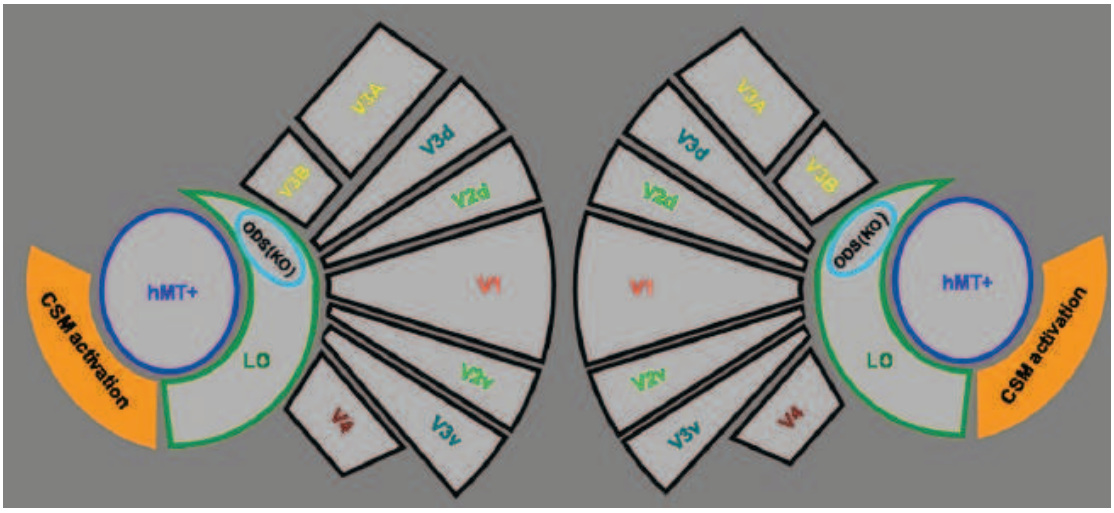


Figure V.3 | A canonical scheme of the typical location of stereomotion activation relative to the established retinotopic and functional regions in occipital cortex: V1–V4, hMT+, LO, ODS (KO). (reproduced from Likova & Tyler, 2007)

Following Likova & Tyler's work, Rokers Cormack & Huk (2009) conducted an fMRI study to map sensitivity to both CDOT and IOVD in the human cortex. In a first experiment, they compared activation for dichoptic (dots going in opposite directions are presented to each eye) and monocular (pairs of dots going in opposite directions are presented to the same eye) displays for horizontal and vertical motion. They observed that activation in MT+ was significantly larger for dichoptic compared to monocular stimuli only for horizontal and not vertical motion, suggesting that these regions are selectively sensitive to motion-in-depth. In a second experiment, they isolated the CDOT component and showed a clear selective activation of areas MT+, V3A and LO. In a third experiment, they annihilated CDOT information by presenting anticorrelated random dots and, again, found an activation of MT+, suggesting that MT+ is sensitive to *both* CDOT and IOVD. Finally, the authors demonstrated the existence of direction-selective adaptation to 3D motion in MT+, consistent with previous psychophysical studies using motion aftereffects (Brooks & Stone, 2004; Fernandez & Farell, 2005; Shioiri et al., 2009). In summary, this fMRI study strongly suggests that static disparity, monocular motion and motion-in-depth are processed in the common area MT+.

1.4 Unresolved questions

1.4.1 Where does the interocular velocity difference come from?

During the last decade, enormous progress has been made on understanding the role of the IOVD cue in the computation of motion-in-depth. For example, it has been demonstrated that IOVD can be especially effective to compute motion-in-depth outside the fovea (Czuba et al., 2010). More importantly, several studies (Brooks & Stone, 2004; Harris & Watamaniuk, 1995; Rokers et al., 2011) have shown that IOVD results from a complex combination of monocular motion signals and that this computation takes place in the area MT+ of the human brain (Rokers et al., 2009).

However, little is known about the *type of calculation* underlying the combination of monocular motion signals and how the *eye-of-origin information* is carried throughout the visual hierarchy to be incorporated in this combination process.

1.4.2 What is motion-in-depth information used for?

Several imaging (Neri et al., 2004) and psychophysical studies (Erkelens & Collewijn, 1985) have suggested the idea that relative and absolute disparity are used in different situations and are represented differently in the brain. While relative disparity would be processed mainly in the ventral stream and used for analysing the 3D shape of objects, absolute disparity would be used through the dorsal stream for orientation and action.

However, evidence concerning the use of motion-in-depth information beyond visual cortical areas is conflicting. Imaging studies mentioned above showed that motion-in-depth is computed in the area MT+ (and anterior to MT+) which is incorporated into the dorsal stream. In addition, several psychophysical studies have reported that sensitivity to motion-in-depth is more similar to the sensitivity of coarse rather than fine stereopsis (Brooks & Stone, 2006). This body of evidence points toward a utility of motion-in-depth information for navigation and action. However, Harris & Sumnall (2000) found that detection of motion-in-depth did not depend on the viewing distance. This result suggests that the computation of motion-in-depth is

not sensitive to the absolute disparity and thus cannot inform on the distance between the observer and the moving object.

1.4.3 How does the visual system keep track of the *change of disparity over time*?

Recent research on motion-in-depth has focused on understanding where and how interocular velocity differences are processed and little is known about the mechanisms underlying the computation of changes of disparity over time.

Using a visual search paradigm, Harris, McKee & Watamaniuk (1998) showed that the detection of a motion-in-depth was more affected by disparity noise than was lateral motion. These authors suggested that the *detection* of 3D motion was carried out by *static* disparity mechanisms rather than specific mechanisms sensitive to the *change* of disparity over time. In Harris & Sumnall's (2000) visual search study, there was no effect of the viewing distance on the detection of 3D and 2D motion, suggesting that motion-in-depth detection is based on *retinal* and not absolute signals.

Likova & Tyler's (2007) imaging study revealed that CDOT information is processed in a specific visual area, anterior to the hMT+ complex and Rokers Cormack & Huk (2009) reported specific activation of V3A and LO regions after presentation of stimuli containing only CDOT information. These two studies thus suggest that static disparity and CDOT are processed in different visual areas.

To investigate how the visual system keeps track of the change of disparity over time, we conducted two series of experiments to examine the temporal and spatial aspects of the computation of 2D and 3D motion. These experiments are presented in the form of two articles in the following sections.

2 Synchronized audio-visual transients drive efficient visual search for motion-in-depth

OPEN ACCESS Freely available online

PLoS one

Synchronized Audio-Visual Transients Drive Efficient Visual Search for Motion-in-Depth

Marina Zannoli[1,2]*, John Cass[3], Pascal Mamassian[1,2], David Alais[4]

1 Université Paris Descartes, Sorbonne Paris Cité, Paris, France, **2** Laboratoire Psychologie de la Perception, CNRS UMR 8158, Paris, France, **3** School of Psychology, University of Western Sydney, Sydney, New South Wales, Australia, **4** School of Psychology, University of Sydney, Sydney, New South Wales, Australia

Abstract

In natural audio-visual environments, a change in depth is usually correlated with a change in loudness. In the present study, we investigated whether correlating changes in disparity and loudness would provide a functional advantage in binding disparity and sound amplitude in a visual search paradigm. To test this hypothesis, we used a method similar to that used by van der Burg et al. to show that non-spatial transient (square-wave) modulations of loudness can drastically improve spatial visual search for a correlated luminance modulation. We used dynamic random-dot stereogram displays to produce pure disparity modulations. Target and distractors were small disparity-defined squares (either 6 or 10 in total). Each square moved back and forth in depth in front of the background plane at different phases. The target's depth modulation was synchronized with an amplitude-modulated auditory tone. Visual and auditory modulations were always congruent (both sine-wave or square-wave). In a speeded search task, five observers were asked to identify the target as quickly as possible. Results show a significant improvement in visual search times in the square-wave condition compared to the sine condition, suggesting that transient auditory information can efficiently drive visual search in the disparity domain. In a second experiment, participants performed the same task in the absence of sound and showed a clear set-size effect in both modulation conditions. In a third experiment, we correlated the sound with a distractor instead of the target. This produced longer search times, indicating that the correlation is not easily ignored.

Citation: Zannoli M, Cass J, Mamassian P, Alais D (2012) Synchronized Audio-Visual Transients Drive Efficient Visual Search for Motion-in-Depth. PLoS ONE 7(5): e37190. doi:10.1371/journal.pone.0037190

Editor: Luis M. Martinez, CSIC-Univ Miguel Hernandez, Spain

Received March 1, 2012; **Accepted** April 18, 2012; **Published** May 17, 2012

Funding: This work was supported by a grant from the French Ministère de l'Enseignement Supérieur et de la Recherche and by a travel grant from Université Paris Descartes. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: marinazannoli@gmail.com

Introduction

For the last fifty years [1], visual search paradigms have proven to be a useful tool to study feature integration [2] and allocation of attention [3]. A majority of studies using this paradigm have focused on the processing of basic feature dimensions such as luminance, color, orientation or motion, and have shown that searching for a target which is distinguished from the surrounding distractors by having, for example, a different orientation (or color, or luminance, etc) produces fast, efficient searches. Most visual search studies employ 2D arrays and relatively few have examined visual search in the 3D domain. Of these, an early study by Nakayama & Silverman [4] showed that distinguishing targets and distractors by their horizontal binocular disparity (stereopsis) was sufficient to support efficient visual search. Later, Harris, McKee & Watamaniuk [5] found that when binocular disparity was defined by spatiotemporal correlations (i.e., perceptual stereomotion), search performance became far less efficient. That is, stereomotion did not support pop-out. This is an intriguing result because even though static stereopsis and stereomotion are each capable of supporting vivid and clearly discriminable perceptual structure, stereomotion seems to require serial search.

In the present study, we will investigate whether search efficiency for stimuli defined by stereomotion can be improved by a non-spatial auditory cue correlated with the visual target. The ability of auditory signals to improve visual processing is now well known. Several studies have shown that the presentation of a simultaneous sound can improve visual performance for detection [6] can increase the saliency of visual events [7] and can drive visual attention [8]. More specifically, using the visual search paradigm, van der Burg and colleagues recently conducted a series of studies on the so-called "pip and pop" effect and demonstrated that a synchronized, but spatially nonspecific, sound can drastically improve search efficiency as long as the visual signal is temporally abrupt [9–11]. In the so-called "pip and pop" effect, search times are drastically decreased for visual objects that are synchronized with an auditory beep even though the sound contains no spatial or identity information concerning the visual target. According to van der Burg and colleagues the auditory "pip" and the visual target are integrated, creating a salient audiovisual object that draws exogenous attention. To test the effect of an auditory cue on visual search for stereomotion stimuli, we used a method similar to the one introduced by van der Burg et al. [10].

The study by van der Burg et al. [10] demonstrated that non-spatial modulations of loudness can drastically improve spatial visual search for a correlated *luminance* modulation but that it requires *transient* visual events (square modulations instead of sine) to elicit efficient search. To enable a comparison with the findings of Van der Burg, et al. [10] in the luminance domain, we decided to use similar modulation conditions. Our participants were

presented with a dynamic random dot stereogram [12] in which 6 or 10 disparity-defined squares arranged on a ring moved back and forth in depth in front of the background plane. Critically, elements in these displays are invisible when viewed monocularly, and require binocular integration across multiple frames. All the elements followed the same spatio-temporal modulation frequency but with different phases. An amplitude-modulating auditory beep was synchronized with the on of the elements' depth modulation. Following the lead of van der Burg, et al. [9,10] we employed a compound search task in which participants performed a discrimination task on a luminance-defined target. The discrimination task is unrelated to the stereomotion but does require participants to successfully find the sound synchronized visual element first.

Although our study uses similar experimental conditions to van der Burg et al. [2], different predictions can be made concerning the modulation conditions. In their study, search for luminance-defined targets was more efficient in the square-wave condition. In our experiment, because binocular matching processes are known to favor smooth over abrupt changes of disparity across space and time [13–15], we predict that the square-modulation condition will not suit stereo processing and will therefore lead to longer response times compared to the sine-modulation condition. In addition, we predict that the presence of the auditory cue will enhance search efficiency in the sine condition and produce smaller set-size effects.

Materials and Methods

Experiment 1

In the first experiment, we tested whether correlating changes in disparity and loudness would provide a functional advantage in binding disparity and sound amplitude in a visual search task. For this purpose, we used visual stimuli moving in depth together with an amplitude-modulating auditory sound with a static location. Participants had to perform a search and a spatial discrimination task on a small 2×2 pixel square defined by luminance. Participants were informed that this luminance target was adjacent to the visual element that was correlated with the accompanying sound changes.

Participants. Five observers (two naïve) with normal or corrected-to-normal vision were recruited in the laboratory building. All participants had experience in psychophysical observation and had normal stereo acuity and hearing. They all gave written informed consent before participating in the experiment.

Stimulus presentation. The stereograms were presented on a 21″ CRT monitor (Sony Multiscan G500, resolution 1024×768 pixels x 85 Hz, for four observers and ViewSonic 2100, resolution 1280×960×85 Hz for one observer) at a simulated distance of 57 cm. To avoid the issues raised by shutter or polarized glasses [16] we used a modified Wheatstone stereoscope. In this type of display, the images presented to the two eyes are completely independent and are presented in perfect synchrony. Each eye viewed one horizontal half of the CRT screen. A chin rest was used to stabilize the observer's head and to control the viewing distance. The display was the only source of light and the stereoscope was calibrated geometrically to account for each participant's interocular distance. The auditory stimuli were presented via a single loudspeaker, which was placed above the monitor.

Stimuli. Stereomotion can be extracted by computing interocular velocity differences and/or by tracking changes of disparity over time [12,17]. In the first case, 2D motion is extracted for each monocular image and then compared between the two eyes' images to compute speed and direction of motion. To avoid any 2D motion cues in the monocular components, we used dynamic random dot stereograms (DRDS). In DRDSs, the stereogram is rebuilt on each new video frame using a new pattern of random noise. Disparity is achieved by adding opposite disparity offsets to a small portion of the left and right images. Stereomotion is then obtained by smoothly changing the value of the disparity offsets from frame to frame. This way, stereomotion in our stimuli was entirely defined by changes of disparity over time. All Stimuli were generated using the Psychophysics Toolbox [18,19].

The background consisted of a 3.5×3.5 deg^2 square of dynamic random noise (mean luminance 40 cd/m^2; one-pixel resolution; refreshed every frame). Visual elements were 0.8×0.8 deg^2 squares defined only by disparity and evenly presented on a virtual ring at 2.5 deg eccentricity. The number of elements was either 6 or 10. A small bright square (2×2 pixels, 80 cd/m^2), too small to capture exogenous attention, was placed either above or below the sound synchronized disparity-defined square to enable a compound search task (see Procedure, below). The background was surrounded by a vergence-stabilization frame consisting of multiple luminance-defined squares (0.20×0.20 deg^2; grey: 40 cd/m^2 and white: 80 cd/m^2) presented on a black background (5 cd/m^2), with black nonius lines at the center (see Figure 1).

Visual elements moved in depth back and forth from 0 to +12 arcmin following a 0.7 Hz modulation. All elements moved at different phases. One of the squares' depth modulation was synchronized with the sound amplitude modulation. To avoid overlapping temporal synchrony between the sound synchronized square and the other visual elements, we created an exclusion

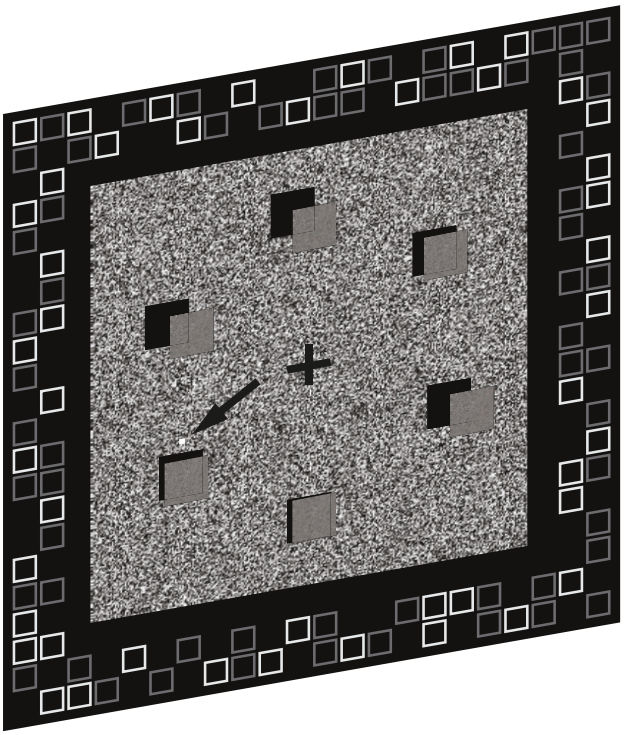


Figure 1. Perspective view of the stimulus used in all experiments. Visual elements were disparity-defined squares distributed evenly on a ring at 2.5 deg eccentricity and moved back and forth in depth from zero to +12 arcmin (crossed) disparity. The stimuli were surrounded by a vergence-stabilisation frame.
doi:10.1371/journal.pone.0037190.g001

window of at least 60° around the sound synchronized square phase: for the other elements, phases were randomly assigned from the following values: ±60°, 80°, 100°, 120°, 140°, 160°, relative to the sound synchronized square's phase.

The auditory stimulus was a 500 Hz sine-wave (44.1 kHz sample rate; mono) whose volume was modulated in amplitude (between 0 and 70 dB) at the same frequency as the visual motion-in-depth and synchronized with the square adjacent to the luminance target. The sound was presented over one loudspeaker placed on top of the CRT screen.

Both visual and auditory modulations were either sine-wave or square-wave and always congruent. A random phase was added to all modulations (see Figure 2). The auditory modulation was synchronized with the depth modulation of the disparity-defined square that was adjacent to the luminance target of the visual search.

Procedure. Participants were instructed to respond as fast as they could while maintaining good performance. Each trial started with a presentation of the nonius lines. When correctly fusing the nonius, participants pressed any key to start the stimulus presentation. In a speeded resnse task, the stimulus stayed on until participants had found the sound synchronized square and made the up/down judgment about the luminance target location and entered their answer on the keypad (which terminated the display). This up/down task (discriminating the position of the luminance target relative to the sound synchronized square) was orthogonal to the stereomotion search (locating the sound synchronized square), as it did not depend on the motion itself. However, as the luminance target was hardly visible while fixating centrally, the localization of the sound synchronized square was necessary first, before the up/down task could be done. This ensured that participants did perceive the disparity-defined squares.

Each combination of waveform condition (square vs. sine) and set size (6 vs 10) was repeated 80 times in total. The experiment was divided in ten sessions. Participants did not receive feedback regarding their accuracy, although they were aware that the amplitude modulation of the auditory signal was synchronized with the visual depth modulation of the adjacent square.

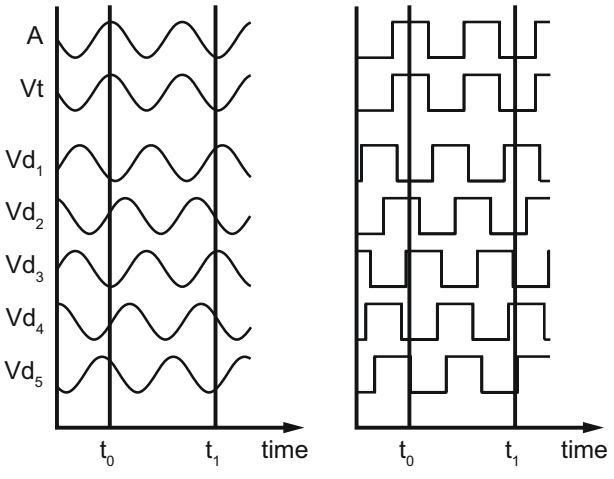


Figure 2. Audiovisual modulations. The depth modulation of the square adjacent to the luminance target is synchronized with an amplitude-modulated 500 Hz tone. Auditory and visual modulations are always congruent (both sine-wave or square-wave). A random phase is added to the AV modulation.
doi:10.1371/journal.pone.0037190.g002

Experiment 2

To test whether results obtained in Experiment 1 are due to the presence of a sound, we tested whether visual sine- and square-wave modulations would lead to different set-size effects in the absence of a congruent auditory modulation.

Method. For the second experiment, the five observers who participated in Experiment 1 (two of whom were naïve) were recruited for Experiment 2. Stimuli were presented using the same setup as in Experiment 1 and the stimuli were identical to the ones used in the first experiment. No auditory signal was presented. Visual elements moved in depth following the same modulation patterns as in Experiment 1. Instructions given to participants were identical to those in Experiment 1.

Experiment 3

In the third experiment, we investigated whether observers were using a voluntary or automatic binding of audiovisual information. We tested this by measuring whether correlating the sound with a square that is not adjacent to the luminance target would lead to longer response times, using a cost-benefit paradigm similar to the one introduced by Posner [3]. In the cost-benefit paradigm, the subject has to perform a discrimination task on a target presented at different locations. Before the presentation of the target stimulus a cue is displayed briefly, indicating the location of the target for that trial. Posner demonstrated that presenting a valid cue (indicating the actual target location) led to shorter response times (i.e., a benefit), relative to a neutral cue (not indicative). On the contrary, presentation of an invalid cue (indicating a wrong location for the target) led to longer response times (i.e., a search cost).

We implemented a cost-benefit experiment in which the square-wave sound could be presented in synchrony with either the square adjacent to the luminance target or another square. 20% of trials were valid (i.e., the sound was synchronized with the adjacent square) and the remaining 80% were invalid trials (i.e., the sound was synchronized with one of the other squares). In invalid trials, if observers were automatically binding the auditory and visual information and going directly to the location where they were synchronized, they would be at a wrong location and would not find the small square there for the up/down discrimination task. They would then have to make a serial search around the depth-modulating visual squares until the one with the small square adjacent to it was found. For this reason, there would be a search cost for invalid if binding were automatic. Alternatively, if the binding of the sound and stereomotion signals were a voluntary strategy, it would be more strategic to ignore the audiovisual correlation (which would be beneficial in only 20% of trials) and begin each trial immediately with a serial search for the small square. If we observe a search cost in the invalid trials (i.e., a slowing of search times), it would show that audiovisual binding was automatic and difficult to ignore.

Method. The five observers who participated in the first two experiments were recruited for the third experiment. Stimuli were presented using the same setup as in the first two experiments. Visual stimuli consisted of nine elements (squares of 0.8×0.8 deg^2) evenly distributed on a ring as in the first two experiments. Auditory stimuli were the same as in Experiment 1. Audiovisual modulations were similar to those in the first experiment (square vs. sine) except that the auditory signal was synchronized with the square adjacent to the luminance target modulation in only 20% of trials. In the remaining 80%, the sound was synchronized with one of the other eight squares. Instructions given to participants were identical as in the first two experiments.

Results

Experiment 1

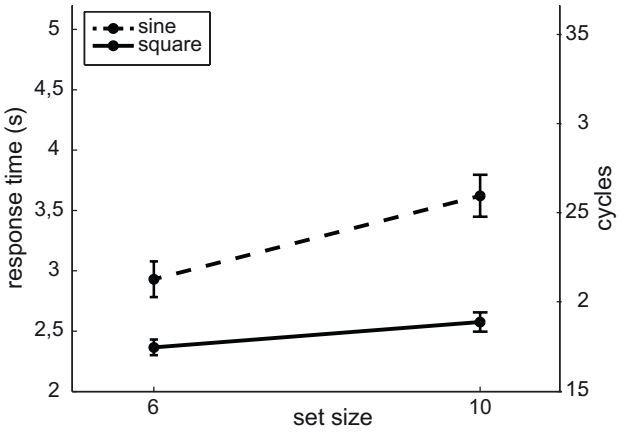
Participants reported that they first localized the sound synchronized square and then saccaded to it to make the up/down judgment concerning the luminance target.

Overall mean error rate was approximately 5% and error trials were discarded and no further analysis was conducted on those data. A cut-off was applied at two standard deviations from the mean response time for each participant (see Figure 3a and 4 and Table S1). A repeated-measures ANOVA was run on the response times with set size (6 vs. 10) and waveform (sine-wave vs. square-wave) as within-subject variables. The ANOVA revealed a significant main effects of set size ($F(1, 3) = 25.9$, $P<0.01$) and waveform ($F(1, 3) = 15.7$, $P<0.05$) and a significant interaction (set size x waveform) effect ($F(3, 1) = 11.6$, $P<0.05$).

Preliminary discussion. As shown in Figure 3a, the significant main effect of waveform arose because response times were faster in the square-wave condition overall. Interestingly, the set size effect was also reduced in the square-wave condition relative to the sine-wave condition. This indicates, contrary to our expectations, that visual search was faster and more efficient in the square wave condition.

In their 2010 study, van der Burg et al. [10] interleaved audiovisual trials with silent trials. This allowed them to interpret the set size effects observed in the audiovisual condition compared to the vision-only trials. During pilot experiments, our participants reported using two distinct conscious strategies depending on whether they were presented an audiovisual or a visual-only trial. Observers would wait for the sound to start to decide which strategy to use. In the presence of a visual-only trial, they would start serial searching for the luminance target while in the case of an audiovisual trial they would maintain central fixation and wait for the synchronized sound square to pop out. If observers were using distinct strategies depending on the condition, it seemed hazardous to compare data collected in the same experiment for these two sets of stimuli.

A Results of Experiment 1



B Results of Experiment 2

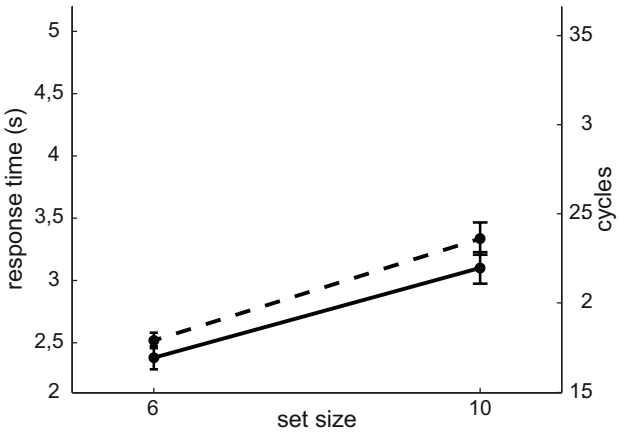


Figure 3. Results of Experiment 1 & 2. Mean response times pooled across five participants as a function of set size and waveform for Experiments 1 (a) & 2 (b). The y-axis on the right represents response times in number of cycles (at 0.7 Hz, 1 cycle lasts 1.4 s). The error bars reflect the overall standard errors of individuals' mean response times. Dashed lines and solid lines code for sine-wave and square-wave modulations respectively.
doi:10.1371/journal.pone.0037190.g003

Experiment 2

If the absence of a set-size effect observed in the square-wave condition in Experiment 1 were due to the auditory information, we expect no difference between the two modulation conditions in the absence of sound. If results from Experiment 2 are comparable to those obtained in Experiment 1, they might reflect a difference in task difficulty between the two modulation conditions. If the square-wave condition is very easy, we might observe a kind of "pop out" effect.

As in Experiment 1, overall mean error rate was approximately 5% and error trials were discarded. A cut-off was applied at two standard-deviations from the mean response time for each participant (see Figure 3b and 4 and Table S2). A repeated-measures ANOVA was run on the response times with set size (6 vs. 10) and waveform (sine-wave vs. square-wave) as within-subject variables. The ANOVA revealed only a significant main effect of the set size ($F(1, 3) = 15.9$, $P<0.05$), with no effect of the waveform ($F(1, 3) = 2.26$, $P = 0.207$) and no significant interaction (set size x waveform) effect ($F(3, 1) = 0.133$, $P = 0.733$). The set-size effect is plotted in Figure 3b. The small difference between the sine- and square-wave conditions is not significant.

Preliminary discussion. In the Experiment 2, we found no significant difference between the two modulation conditions. Both sine- and square-wave conditions led to significant and comparable set-size effects. This confirms that the absence of a set-size effect in the square modulation condition of Experiment 1 can be attributed to the synchronized presence of a transient auditory signal. In addition, participants responded more quickly on the visual search task in Experiment 2 than in Experiment 1. This effect could be explained by participants using distinct conscious strategies for audiovisual and visual-only trials, as suggested in the Discussion of Experiment 1. If so, the facilitation in visual search observed in the square-wave condition of Experiment 1 could be due to a voluntary binding of visual and auditory information. To test this assumption, we used a cost–benefit paradigm in Experiment 3.

Experiment 3

As in the first two experiments, overall mean error rate was approximately 5% and error trials were discarded. A cut-off was applied at 2 standard-deviations from the mean response time for each participant (see Figure 5a and 5b and Table S3). A repeated-measures ANOVA was run on the response times with cue validity (valid vs. invalid) and waveform (sine-wave vs. square-wave) as within-subject variables. The ANOVA revealed a significant effect of cue validity ($F(1, 3) = 15.3$, $P<0.05$), no effect of the waveform

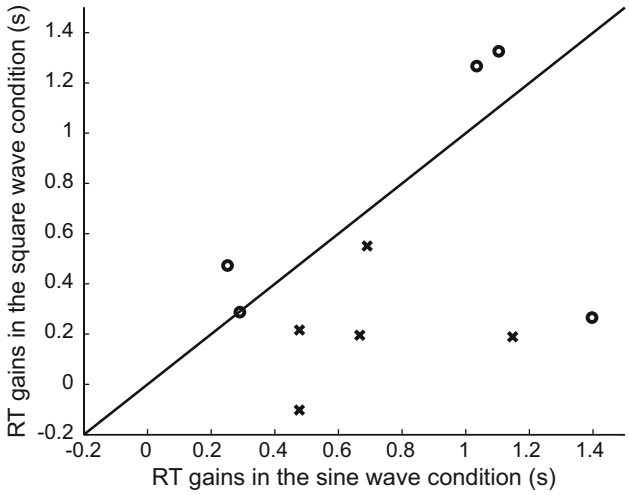


Figure 4. Individual results of Experiment 1 & 2. Response time (RT) gains ((RT(10) - RT(6)) in the square-wave condition as a function of the response time gains in the sine-wave condition. Along the black line, slopes are equal for both waveforms. When individual points are located in the lower part of the figure, response time gains are smaller in the square-wave condition. Crosses and dots represent individual results in Experiment 1 & 2 respectively.
doi:10.1371/journal.pone.0037190.g004

($F(1, 3) = 2.84$, $P = 0.167$) and a significant interaction (cue validity * waveform) effect ($F(3, 1) = 8.47$, $P < 0.05$).

Preliminary discussion. The results of Experiment 3 (Figure 5a) show a clear benefit in the square- compared to the sine-wave condition when the sound was synchronized with the adjacent square, and a cost when the square-wave sound was synchronized with one of the other squares. Even though the sound correlated with the adjacent square in only 20% of the trials, which all observers knew, results suggest that observers were unable to stop using the audiovisual synchrony. In 80% of trials, this strategy led to a wrong square and consequently slowed down the visual search process. This cost effect implies that the audio-visual correlation was automatically bound and could not be easily ignored.

Discussion

The goal of this series of experiments was to explore the effect of an auditory cue on visual search for stereomotion-defined visual stimuli. In the first two experiments, we showed that an amplitude-modulating auditory beep synchronized with a visual target led to efficient visual search. On the face of it, this result seems to contradict the finding from Harris, et al. [5] that stereomotion does not pop out. Moreover, we found a significant improvement in visual search only when the auditory and visual modulations were square and not sine. Our results add to those obtained by van der Burg et al. [10] by showing that pip and pop is neither the exclusive domain of the luminance system, nor is it purely monocularly-driven.

Our predictions were that, contrary to the luminance system, the stereo system would be more efficient at tracking smooth (sine-wave) rather than abrupt (square-wave) changes of disparity over time. Instead, we found that visual search was more efficient for square-wave than for sine-wave modulations of depth. This suggests that the stereo system is better able to keep track of rapid temporal modulations in spatio-temporal disparity when guided by an auditory cue.

The third experiment was aimed at investigating whether the results from Experiments 1 and 2 could be attributed to an

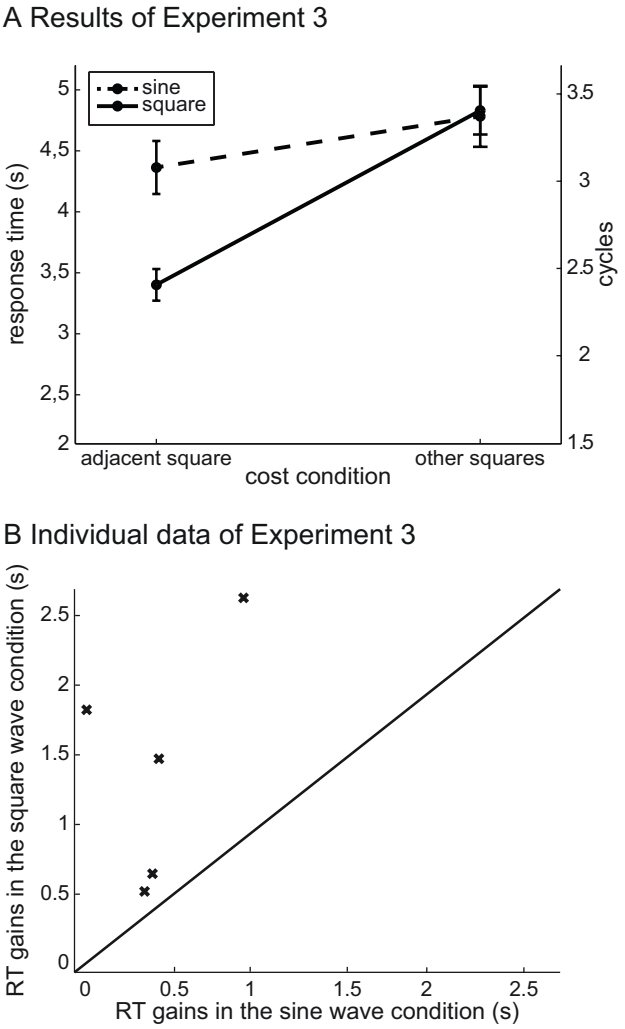


Figure 5. Results of Experiment 3. (A) Mean response times pooled across five participants as a function of cue validity and waveform. See legend from Figure 3 for details. (B) Individual results of Experiment 3. Response time gains (RT(other squares) - RT(adjacent square)) in the square-wave condition as a function of the response time gains in the sine-wave condition. See legend from Figure 4 for details.
doi:10.1371/journal.pone.0037190.g005

automatic integration of auditory and visual temporal signals or to a voluntary attention-like effect. The results of this last experiment suggest that even when the sound led to wrong locations and thus impaired visual search, the correlation between the auditory and visual signals could not be easily ignored. This conclusion is consistent with an interpretation in terms of audiovisual integration rather than one of crossmodal attention.

Neural structures differentially responsive to synchronized audiovisual events have been found throughout the human cortex [7]. Recently, luminance-driven pip and pop-related increases in event related potentials were observed over lateral occipital areas of cortex [11]. It is conceivable that the compulsory audio-visual integration we observe may be related to audio-visually evoked activity in similar cortical areas.

The results of the experiments described in this article suggest that three main conclusions. First, an auditory cue can significantly improve the detection of targets defined exclusively by stereomotion, and second, that the stereo system is able to track abrupt

changes of disparity over time when it is paired with a synchronized auditory signal. Third, and more generally, our findings support the idea that the pip and pop effect is likely to be mediated at a cortical level as we have demonstrated it here with stimuli that are exclusively binocularly defined.

Supporting Information

Table S1 Individual data of Experiment 1. Individual response times (s) as a function of set size and waveform for Experiment 1.
(DOCX)

Table S2 Individual data of Experiment 2. Individual response times (s) as a function of set size and waveform for Experiment 2.
(DOCX)

Table S3 Individual data of Experiment 3. Individual response times (s) as a function of cue validity and waveform for Experiment 3.
(DOCX)

Acknowledgments

This work was first presented at the annual conference of the Vision Sciences Society in Naples, Florida, in May 2011 [20]. We thank Laurie Wilcox for discussions. The experiments were approved by the Ethics Committee of the Université Paris Descartes.

Author Contributions

Conceived and designed the experiments: MZ JC PM DA. Performed the experiments: MZ. Analyzed the data: MZ. Wrote the paper: MZ JC PM DA.

References

1. Neisser U (1964) Visual search. Scientific American.
2. Treisman AM, Gelade G (1980) A feature-integration theory of attention. Cognitive Psychology 12: 97–136.
3. Posner MI (1980) Orienting of attention. Quarterly Journal of Experimental Psychology 32: 3–25.
4. Nakayama K, Silverman GH (1986) Serial and parallel processing of visual feature conjunctions. Nature 320: 264–265.
5. Harris JM, McKee SP, Watamaniuk SN (1998) Visual search for motion-in-depth: stereomotion does not "pop out" from disparity noise. Nature Neuroscience 1: 165–168.
6. Andersen TS, Mamassian P (2008) Audiovisual integration of stimulus transients. Vision Research 48: 2537–2544.
7. Noesselt T, Tyll S, Boehler CN, Budinger E, Heinze HJ, et al. (2010) Sound-Induced Enhancement of Low-Intensity Vision: Multisensory Influences on Human Sensory-Specific Cortices and Thalamic Bodies Relate to Perceptual Enhancement of Visual Detection Sensitivity. Journal of Neuroscience 30: 13609–13623.
8. Lippert M, Logothetis NK, Kayser C (2007) Improvement of visual contrast detection by a simultaneous sound. Brain Research 1173: 102–109.
9. van der Burg E, Olivers CNL, Bronkhorst AW, Theeuwes J (2008) Pip and pop: nonspatial auditory signals improve spatial visual search. Journal of Experimental Psychology: Human Perception and Performance 34: 1053–1065.
10. van der Burg E, Cass J, Olivers CNL, Theeuwes J, Alais D (2010) Efficient Visual Search from Synchronized Auditory Signals Requires Transient Audiovisual Events. PLoS ONE 5: e10664.
11. van der Burg E, Talsma D, Olivers CNL, Hickey C, Theeuwes J (2011) Early multisensory interactions affect the competition among multiple visual objects. NeuroImage 55: 1208–1218.
12. Harris JM, Nefs HT, Grafton CE (2008) Binocular vision and motion-in-depth. Spatial vision 21: 531–547.
13. Marr D, Poggio T (1976) Cooperative computation of stereo disparity. Science 194: 283–287.
14. Assee A, Qian N (2007) Solving da Vinci stereopsis with depth-edge-selective V2 cells. Vision Research 47: 2585–2602.
15. Nienborg H, Bridge H, Parker AJ, Cumming BG (2005) Neuronal computation of disparity in V1 limits temporal resolution for detecting disparity modulation. Journal of Neuroscience 25: 10207–10219.
16. Tsirlin I, Wilcox LM, Allison RS (2011) The effect of crosstalk on the perceived depth from disparity and monocular occlusions. Broadcasting, IEEE Transactions on. pp 1–9.
17. Brooks KR, Stone LS (2006) Spatial scale of stereomotion speed processing. Journal of Vision 6: 9–9.
18. Brainard DH (1997) The Psychophysics Toolbox. Spatial vision 10: 433–436.
19. Pelli DG (1997) The VideoToolbox software for visual psychophysics: transforming numbers into movies. Spatial vision 10: 437–442.
20. Zannoli M, Cass J, Mamassian P, Alais D (2011) Synchronized audio-visual transients drive efficient visual search for motion-in-depth. Journal of Vision 1: 11. 792 p.

3 Stereomotion detectors are poorly suited to track 2D motion

Disparity-based stereomotion detectors are poorly suited to track 2D motion

Marina Zannoli — Université Paris Descartes, Sorbonne Paris Cité, Paris, France; Laboratoire Psychologie de la Perception, Paris, France

J. Cass — School of Psychology, University of Western Sydney, Sydney, New South Wales, Australia

D. Alais — School of Psychology, University of Sydney, Sydney, New South Wales, Australia

P. Mamassian — Université Paris Descartes, Sorbonne Paris Cité, Paris, France; Laboratoire Psychologie de la Perception, Paris, France

?1 A study was conducted to examine the time required to process lateral motion and motion-in-depth for luminance- and
?2 disparity-defined stimuli. In a 2 × 2 design, visual stimuli oscillated sinusoidally in either 2D (moving left to right at a constant
?3 disparity of 9 arcmin) or 3D (looming and receding in depth between 6 and 12 arcmin) and were defined either purely by
?4 disparity (*change of disparity over time* [CDOT]) or by a combination of disparity and luminance (providing CDOT and
?5 *interocular velocity differences* [IOVD]). Visual stimuli were accompanied by an amplitude-modulated auditory tone that
?6 oscillated at the same rate and whose phase was varied to find the latency producing synchronous perception of the auditory and visual oscillations. In separate sessions, oscillations of 0.7 and 1.4 Hz were compared. For the combined CDOT + IOVD stimuli (DL conditions), audiovisual synchrony required a 50 ms auditory lag, regardless of whether the motion was 2D or 3D. For the CDOT-only stimuli (DO conditions), we found that a similar lag (~60 ms) was needed to
?7 produce synchrony for the 3D motion condition. However, when the CDOT-only stimuli oscillated along a 2D path, the auditory lags required for audiovisual synchrony were much longer: 170 ms for the 0.7 Hz condition, and 90 ms for the 1.4 Hz condition. These results suggest that stereomotion detectors based on CDOT are well suited to tracking 3D motion, but are poorly suited to tracking 2D motion.

Keywords: stereomotion, stereopsis, motion, audio-visual integration

Citation: Zannoli, M., Cass, J., Alais, D., & Mamassian, P. (2012). Disparity-based stereomotion detectors are poorly suited to 2D motion. *Journal of Vision*, *12*(11):x, xx–xx, http://www.journalofvision.org/content/12/11/x, doi:10.1167/12.11.x.

Introduction

To estimate the depth order relationships between objects, the visual system relies on multiple cues to depth. One such cue is binocular disparity. Stereopsis refers to the perception of depth derived from disparities between the two eyes' retinal images. Whilst static objects may be defined purely by their interocular *spatial* correlations, objects undergoing motion-in-depth are defined by correlations that co-occur across space *and* time. There are two main cues to extract motion-in-depth. The visual system can extract binocular disparity of an object relative to the fixation plane and track how this information varies over time (change of disparity over time [CDOT]). Another possibility is to combine the velocity of an object extracted from each monocular image (interocular velocity difference [IOVD]). Since the seminal work of Rashbass and Westheimer (1961), extensive psychophysical work has been done to understand the nature and the relative utility of these two cues that are often present redundantly in natural scenes (Harris, Nefs, & Grafton, 2008; Nefs & Harris, 2010). A majority of the studies conducted in the last decade have focused on understanding how monocular motion signals are combined to detect motion-in-depth (Cumming & Parker, 1994) and to discriminate the speed of motion-in-depth (Brooks, 2002; Brooks & Stone, 2004; Harris & Watamaniuk, 1995; Rokers, Czuba, Cormack, & Huk, 2011).

Less interest has been shown in understanding the mechanisms underlying the tracking of changes in

doi: 10.1167/XX.XX.XX Received June 24, 2012; published Month 0, 2012 ISSN 1534-7362

disparity over time. Harris, McKee, and Watamaniuk (1998) showed that the detection of motion-in-depth was more affected by disparity noise than was the detection of lateral motion, and they suggested that the detection of 3D motion was carried out by *static* disparity mechanisms rather than specific mechanisms sensitive to the *change* of disparity over time. Using the same paradigm, Harris and Sumnall (2000) showed that there was no effect of the viewing distance on the detection of 3D and 2D motion, suggesting that motion-in-depth detection is based on *relative* and not absolute signals.

To complement these behavioral results, two fMRI studies have proposed that motion-in-depth is computed by specific neurons in the visual cortex. First, Likova and Tyler (2007) recorded bold activation in the dorsal stream after presentation of motion-in-depth stimuli containing only CDOT information and discovered a visual area, anterior to hMT+ (previously found to be sensitive to motion and disparity information; Maunsell & Van Essen, 1983) exclusively sensitive to changes of disparity over time. Later, Rokers, Cormack, and Huk (2009) found that area hMT+ was sensitive to both CDOT and IOVD types of information. In addition, they reported specific activation of V3A and LO after presentation of stimuli containing only CDOT
?8 information. Taken together, these results suggest that: (a) changes of disparity over time are mediated by cortical mechanisms separate from those associated with the processing of static disparity signals; and (b) that these CDOT-selective mechanisms are associated with the perception of motion-in-depth.

At the same time, another line of work focused on understanding the interactions between the processing of motion and binocular disparity. Maunsell and Van Essen (1983) were the first to report the existence of cells sensitive to both binocular disparity and frontoparallel motion in macaque MT, suggesting that lateral
?9 motion is treated separately for different depth planes. This was confirmed by more recent psychophysical work on motion transparency. For example, Hibbard and Bradshaw (1999) and Snowden and Rossiter (1999) measured thresholds for the identification of the direction of motion for stimuli in which signal and noise elements were given various disparities, and they found that performance was substantially better when signal and noise had different disparities. Similarly, Edwards and Greenwood (2005) and Greenwood and Edwards (2006) showed that observers are able to detect a larger number of transparent motion directions when they are carried by signals that are distributed across distinct depth planes.

Even though the processing of motion and binocular disparity seem to share common cortical resources, their underlying mechanisms have different spatial and temporal resolutions. In the stereo domain, both temporal and spatial resolution have been found to be worse than for lateral motion (Norcia & Tyler, 1984; Regan & Beverley, 1973; Tyler, 1971). It has also been shown that differences in temporal resolution and time of processing can be found within the stereo system itself. For example, Julesz (1960) observed that depth from random-dot stereograms (RDSs) took more time than for stimuli with monocular segmentation information. However, more recently Uttal, David, and Welke (1994) reported that observers were above chance when asked to recognize a 3D shape on a RDS presented for 1 ms. Both studies showed an effect ?10 of practice on the latency of stereopsis from RDSs.

The aim of the present study was to investigate how long it takes the visual system to process changes in direction of motion, comparing stimuli oscillating at a constant (nonzero) disparity over time (2D motion in the frontoparallel plane) with stimuli oscillating through varying disparities over time (3D motion-in-depth). Performance in these lateral motion and motion-in-depth conditions are compared for stimuli with and without the contribution of monocular segmenting information. In this way, we compare the temporal resolution of the luminance and disparity-defined motion systems for detecting changes in the direction of moving stimuli in a 2D or 3D context. We do this using a method introduced by Moutoussis and Zeki (1997a, b) to study relative processing latencies between different stimulus attributes. Their stimulus consisted of a pattern of colored squares that oscillated in position (up/down) and color (red/green) following a square-wave pattern. By shifting the phase of the color alternation relative to motion until they both appeared to change synchronously, they showed that color changes were perceived 70–80 ms before motion changes. They then verified that this was a fixed offset by testing different frequencies of color/motion change.

We employ an analogous paradigm to investigate the processing latencies for changes in perceived direction of luminance- and disparity-defined motion within and across depth planes. As this method involves continuously cycling stimuli, it has an important advantage over other paradigms using brief stimuli because the phase-lag required for perceptual synchrony can be assumed to reflect a pure latency difference and not to include the time needed to fuse the two eyes' images and compute the disparity map, which would happen only once, at stimulus onset. Because our measure is free of a time-to-fuse component, it allows us to compare the optimal latency for synchrony with the same measure obtained in the luminance domain, and for lateral motion versus motion-in-depth. Several studies using visual objects defined by luminance have reported that the auditory event must be presented 30–40 ms after the visual stimulus to perceive audiovisual synchrony (Lewald & Guski, 2003). However, little is

known about the time required to compute audiovisual simultaneity for disparity-defined visual stimuli.

Method

Participants

Five observers (4 naïve and 1 author) with normal or corrected-to-normal vision were recruited from the laboratory. All had experience in psychophysical observation and had normal stereo acuity and hearing.

Stimulus presentation

The stereograms were presented on a CRT monitor (ViewSonic 21", resolution of 1280 × 960, refresh rate of 85 Hz) using a modified Wheatstone stereoscope at a ?11 simulated distance of 57 cm. Each eye viewed one horizontal half of the CRT screen. A chin rest was used to stabilize the observer's head and to control the viewing distance. The display was the only source of light and the stereoscope was calibrated geometrically to account for each participant's interocular distance. The auditory stimuli were presented binaurally through headphones.

Stimuli

Visual stimuli

Stimuli were generated using the Psychophysics ?12 Toolbox (Brainard, 1997; Pelli, 1997). To avoid any coherent 2D motion signals in the monocular images of our disparity-defined stereomotion stimuli, we used dynamic random-dot stereograms (DRDS). In DRDSs, the stereogram is rebuilt on each video frame using a new pattern of random noise. Disparity was achieved by adding horizontal offsets in interocular disparity to a small portion of the left- and right-eye images. Stereomotion was then obtained by smoothly changing the value of the disparity offsets from frame to frame. This way, stereomotion in our stimuli was entirely defined by changes of binocular disparity over time.

The background consisted of a 3.1° × 3.1° square of dynamic random noise (mean luminance 40 cd/m^2; one-pixel resolution; refreshed every frame). Visual stimuli (see Figure 1) consisted of fifteen randomly distributed squares (0.6° × 0.6°) defined either by disparity and luminance (DL) conditions or by disparity only (DO) conditions. In the DO condition, the squares were composed of dynamic random noise and were distinguished from the background only by disparity (and therefore visible only binocularly), whereas in the DL condition, they were black squares of 5 cd/m^2 luminance (and therefore monocularly visible), which were also disparate relative to the background. In the DL conditions, both CDOT and IOVD cues were available, while only the CDOT cue was present in the DO conditions. Each square in the set consistently moved from left to right in opposite directions in each eye between 6 and 12 arcmin, producing a percept of motion-in-depth (3D motion), or from left to right (6 arcmin amplitude) at a constant disparity of 9 arcmin, producing a percept of lateral displacement at a pedestal depth (2D motion).

The background was surrounded by a vergence-stabilization frame consisting of multiple luminance-defined squares (0.20° × 0.20°; gray: 40 cd/m^2 and white: 80 cd/m^2) presented on a black background (5 cd/m^2). Black nonius lines were presented at the center of the display (see Figure 1).

Auditory stimuli

The auditory stimulus was a 500 Hz sine-wave (44.1 kHz sample rate; mono) whose envelope (amplitude) was modulated between 0 and 70 dB at the same frequency as the modulations in visual motion direction. The sound was presented binaurally through headphones.

Audiovisual modulations

The audiovisual stimulus was presented for 2 s. In order to test whether the optimal latencies measured were dependant on the phase of the motion (in degrees) or whether they reflected absolute latencies (in ms), the experiment was replicated for two different frequency values: 0.7 (equivalent speed: 0.14°/s) and 1.4 Hz (equivalent speed: 0.28°/s). These values were chosen in order to maximize sensitivity to CDOT (Czuba, Rokers, Huk, & Cormack, 2010; Shioiri, Nakajima, Kakehi, & Yaguchi, 2008).

Both modulations were sinusoidal. The phase of the auditory modulation relative to the visual modulation varied between 0° and 345° in steps of 15°. A random phase was added to all modulations (see Figure 2).

Procedure

The experiment was divided into two sessions. In the first session, the audiovisual modulations were at a frequency of 0.7 Hz. In the second session, the frequency was doubled to 1.4 Hz. For each session, the four conditions of DO/DL * 2D/3D motion were presented in separate blocks. The four blocks were presented in a ?13 random order. Each block contained a total of 192 trials (8 repetitions for each of the 24 auditory phases).

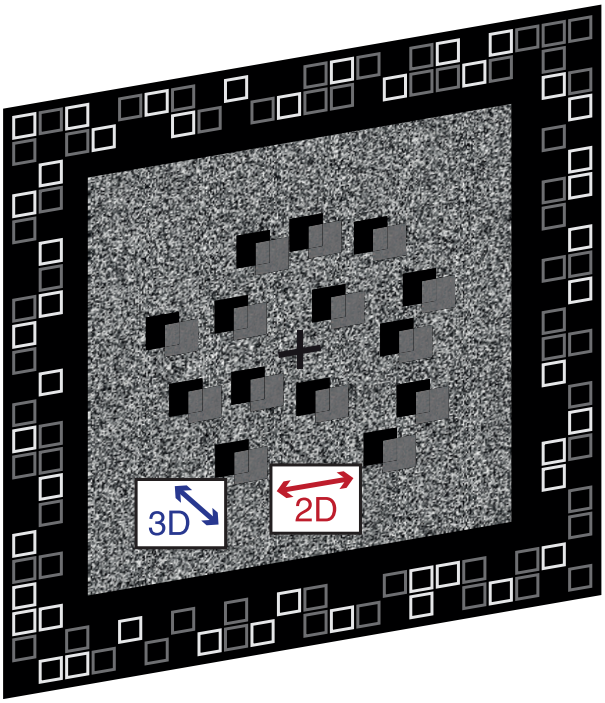


Figure 1. Perspective view of the stimulus. Fifteen squares are randomly located on a dynamic random dot stereogram background. Squares and background are shown in a different resolution for the purpose of the schematic representation. The squares can (1) be defined only by disparity (DO condition) and thus can be seen only when the two eyes' images are fused or (2) be defined by disparity and by luminance (5 cd/m^2; DL condition) and be seen monocularly. The squares follow either a 2D or a 3D motion direction. In either case, the amount of displacement is identical (6 arcmin).

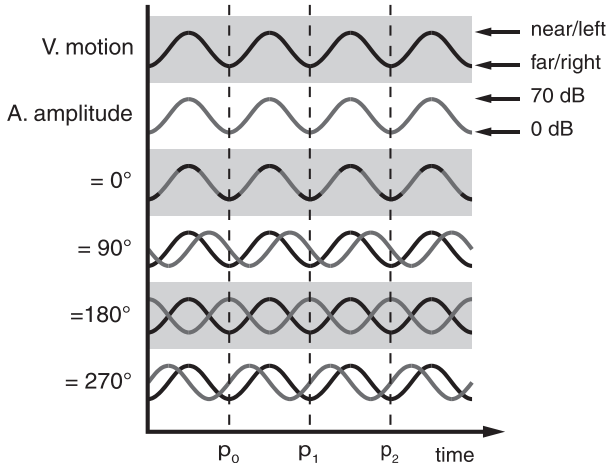


Figure 2. Schematic representation of the audiovisual modulations. The pattern of squares moves back and forth (from 6 to 12 arcmin) or from left to right (6 arcmin amplitude displacement at 9 arcmin disparity), while the auditory signal modulates in amplitude between 0 dB and 70 dB. The phase of the auditory signal relative to the visual modulation is randomly shifted by an angle of 0° to 345° in steps of 15°, inducing either an auditory or a visual lag.

Participants were asked to match the direction of motion with the amplitude modulation. In the 3D motion conditions, participants were asked to press one key of a keyboard if the maximum auditory amplitude was synchronized with the squares being at their perceptually "farthest" position and another key if the maximum of the sound amplitude was synchronized with the squares being at their perceptually "closest" position. In the 2D motion conditions, participants used the left and right arrows to respectively indicate audio-visual synchrony between the maximum amplitude of the tone and left-most and right-most points in the 2D motion trajectory.

Results

Figure 3 shows the averaged response curves for the 5 participants. The proportion of "near" (for the 3D motion condition) or "left" (for the 2D motion condition) responses is plotted as a function of auditory latency. A negative latency represents an auditory signal lag while a positive latency codes for a visual lag. When an audio-visual phase lag of zero is applied to the auditory signal, the maximum amplitude (70 dB) is synchronized with the maximum visual disparity value (12 arcmin), or with the left position for the 2D motion condition. If perception of the auditory and visual changes occurred with no differential latency, we would expect the maximum of the response curve to peak at a value of 0 ms. If this maximum deviates from 0 ms, it suggests that visual and auditory information are perceived at different times. We fitted logit functions (Mamassian & Wallace, 2010; see Figure 3) to the distributions and extracted slopes for the four conditions. For each latency θ the probability p to perceive the maximum of auditory amplitude synchronized with the "left" or "near" position (depending on the motion condition) of the visual stimuli is characterized by the following logit model:

$$\mathrm{logit}(p) = \ln\left(\frac{p}{1-p}\right) = \gamma - \beta_\theta|\theta - \theta_0|_\pi$$

where θ_0 is the optimal latency, β_θ represents the strength of the effect of latency on the proportion of "left" or "near" responses, and γ is a constant. In this equation, $|\ \ |_\pi$ stands for the absolute value modulo π, i.e., $|X|_\pi = \mathrm{acos}(\cos[x])$. The parameter β_θ shows how sensitive an observer is for small variations of latency ?14

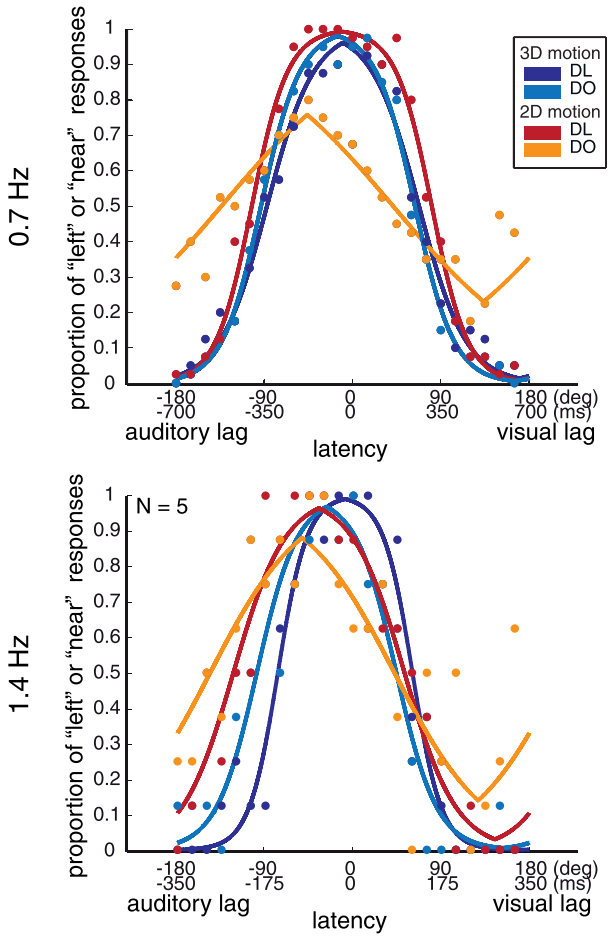


Figure 3. Results of the experiment. This plot shows the proportion of "left" or "near" responses as a function of the latency. The data is pooled across the 5 participants. The visual stimuli were defined by 3D or 2D motion (blue or red) and by DO or by DL (light or dark). A logit function was fitted to the data from the four experimental conditions.

(its unit is in ms^{-1} when latencies are expressed in ms). Figure 4 shows the group mean slopes as a function of the latencies extracted from the best-fitting logit functions for the four conditions at each oscillation rate.

A repeated-measures ANOVA was run on the mean latency and the slope with the type of stimuli (DO vs. DL) and the type of motion (2D vs. 3D) as within-subject variables for the two frequency values sessions (0.7 and 1.4 Hz). The ANOVA for the 0.7 Hz session revealed a significant effect of the type of stimuli, $F(1, 3) = 14.8$, $p < 0.01$ for the slope and $F(1, 3) = 13.6$, $p < 0.05$ for the latency; the type of motion, $F(1, 3) = 12.0$, $p < 0.05$ for the slope and $F(1, 3) = 9.96$, $p < 0.05$ for the latency; and a significant interaction (type of stimuli × type of motion) effect, $F(3, 1) = 90.7$, $p < 0.01$ for the slope and $F(3, 1) = 14.8$, $p < 0.05$ for the latency. For the 1.4 Hz session, the ANOVA revealed a significant effect of the type of stimuli, $F(1, 3) = 66.2$, $p < 0.01$, and the type of motion, $F(1, 3) = 35.5$, $p < 0.01$, but no significant interaction effect, $F(3, 1) = 6.58$, $p = 0.06$ for the slope. For the latency measure in the 1.4 Hz session, the ANOVA revealed no significant effect of the type of stimuli, $F(1, 3) = 3.86$, $p = 0.12$, a significant effect of the type of motion, $F(1, 3) = 19.3$, $p < 0.05$, and a significant interaction effect, $F(3, 1) = 9.74$, $p < 0.05$. To further investigate the effects found in the ANOVAs, we tested multiple comparisons with Tukey least-significant difference corrections. In the 0.7 Hz session, for the slope and latency measures, we found no difference between the DO-3D, DL-2D, and DL-3D conditions and a significant difference between the DO-2D condition and the three other conditions. In the 1.4 Hz session, for the slope measures, we found no difference between the DO-3D, DL-2D, and DL-3D conditions and a significant difference between the DO-2D condition and the three other conditions. For the latency measure, only the comparison between the DO-3D and DO-2D conditions was significant.

A casual exploration of Figure 4 suggests a potential relationship between latency and slope: small latencies (i.e., low bias) are linked to large slopes (i.e., high sensitivity). However, with only eight conditions (and a clear outlier), this apparent relationship should be taken with caution.

The 2D and 3D motion conditions for DL stimuli were similar in terms of optimal latency for perceived synchrony (mean auditory lag: 43 and 37 ms for the 0.7 Hz condition and 59 and 48 ms for the 1.4 Hz condition for the 2D and 3D motion conditions, respectively). Surprisingly, even though stereopsis is often thought to be a slow process, we found the optimal latency for DO-3D motion stimuli was only slightly longer (mean auditory lag: 55 ms and 64 ms for the 0.7 Hz and 1.4 Hz conditions, respectively). However, when participants had to judge synchrony for the DO-2D motion stimuli, it led to larger latencies (170 and 90 ms for the 0.7 Hz and 1.4 Hz conditions). In addition, in the DO-2D motion, the slope of the distribution was substantially shallower than in the three other conditions for the two frequency conditions, suggesting that the task was much harder (see Figures 3 and 4).

We found a similar pattern of results in the two experiments (similar latencies and slopes for three conditions and longer latency and shallower slope in the DO-2D motion condition). Latencies in the two experiments are equivalent in terms of absolute latencies except for the DO-2D motion condition. In this condition, the latency was divided by two in the 1.4 Hz experiment compared to the 0.7 Hz experiment.

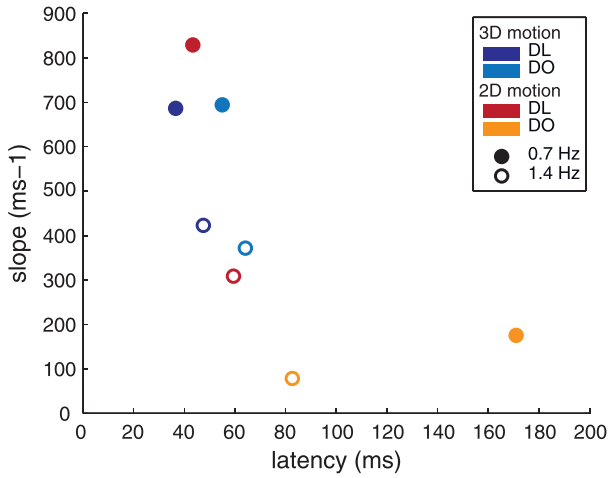


Figure 4. Results of the experiment. Slope as a function of latency for the four experimental conditions in the 0.7 Hz (plain dots) and 1.4 Hz (empty dots) experiments. The mean latency is significantly larger in the DO-2D motion condition than in the three other conditions only for the 0.7 Hz session. The slope is significantly smaller in the DO-2D motion condition than in the three other conditions for the two modulation frequencies.

Discussion

To sum up, we measured the latencies required to perceptually align an auditory modulation with an oscillating visual motion. We compared lateral (2D) motion and motion-in-depth (3D), for motion tokens ?15 defined either by DL, or by DO only. Of these four conditions, we found a similar optimal latency for audiovisual synchrony in three conditions: 2D and 3D motion for the luminance stimuli and 3D motion for the DO condition all produced latencies in the range of 50–60 ms. The exception was in the DO-2D condition where the latency was up to three times larger than for the three other conditions. In addition, the slope of the distribution of synchrony judgments for this particular experimental condition was much shallower than for the other three. Together, these results indicate that the stereomotion system is able to detect changes in the direction of motion as rapidly and precisely as the luminance system can detect direction changes, provided the signal contains changes in depth. Even though this result might seem at odds with previous observations by Tyler (1971), we think that it is hazardous to compare our results with Tyler's because of several empirical differences. Tyler measured movement sensitivity, so he used small motion amplitudes that were difficult to perceive. We measured the optimal latency for the perception of synchrony between sound and visual motion. For this purpose, we used stimuli in which displayed motion (2D or 3D) was suprathreshold. Another difference is that our visual stimuli moved around a disparity pedestal, whereas Tyler's were around zero disparity.

The second main result of our study is, however, that when disparities do not vary across time, as in lateral motion at a fixed nonzero disparity, the stereomotion system is very sluggish.

According to the *continuity* rule stated by Marr and Poggio (1976), smooth modulations of disparity over time are easier to detect than abrupt changes. Let us consider a limited area adjacent to the edge of one of the squares present in the visual stimulus on the DO-2D condition. Through this small window, the edge of the square is successively present or absent creating abrupt changes of disparity. If the stereo system relied on such transient information to compute the 2D motion in this stimulus, the task would be much harder than for the other stimulus configurations, leading to degraded performances. We ran a control experiment to test whether the performance obtained in the DO-2D condition could be explained by the temporal integration of square (on/off) modulations of disparity in a limited area of the stimulus. The same participants ran two separate sessions similar to the ones from the main experiment. For both frequency conditions, the mean slope from the control condition was significantly steeper than for the DO-2D condition, $t(4) = 3.1$, $p < 0.05$ for the 0.7 Hz condition and $t(4) = 3.53$, $p < 0.05$ for the 1.4 Hz condition, suggesting that the task was easier in the control condition. Therefore, degraded performances in the DO-2D condition cannot be accounted for by local integration of square modulations of disparity.

In a pilot experiment run on one author, we also tested whether introducing a small amount of 3D motion (1.2 and 2.4 arcmin) would result in a reduction of latency and an increase in slope. We found that slopes and latencies in these two conditions were similar than in the 2D motion condition.

It is of particular interest to compare the two DO conditions. In these two conditions, the moving squares sustained the same amplitude of motion. In the 3D motion condition, the direction of motion (laterally) was in antiphase in one eye compared to the other, while the direction of motion was identical in the two eyes in the 2D motion condition. Therefore, 2D and 3D motion conditions differed only in the direction of lateral displacement across the eyes. It is likely that this difference is responsible for the optimal latency and performance differences between these two conditions.

Implications of the optimal latency reports

The method employed in the present study has been used in several psychophysical works to assess the

timing of processing different perceptual attributes. Following Moutoussis and Zeki's (1997a, 1997b) work on color and motion, Zeki and Bartels (1998) argued that the activity of neurons in a given system is sufficient to elicit a conscious experience of the attribute that is being processed. Therefore, the optimal latency for the perception of synchrony between two attributes directly reflects the timing of processing of these two attributes. Moutoussis and Zeki (1997a, 1997b) found that color information is perceived 60–80 ms before motion information. Stone et al. (2001) measured the point of subjective simultaneity between a light and a sound using a method similar to Moutoussis and Zeki (1997a, 1997b) and found that this optimal latency measure was observer-specific (ranging from −21ms to +150 ms of auditory lag) and stable. Our results add to these previous observations by showing that the disparity system takes longer to process lateral motion than motion-in-depth.

Implications of the performance measures

The slopes extracted from the logit function fitted to the raw data add to the optimal latency reports and suggest that not only does the disparity system take longer to process changes in the direction of lateral motion than motion-in-depth, but also that it is less efficient at doing so. While it appears from our results that there exists a specific system dedicated to extract motion-in-depth from changes of disparity over time, lateral motion must be inferred from a series of snapshots when moving objects are defined only by disparity.

This result is in contradiction with a basic assumption of a majority of the physiological and computational models of stereopsis. Most cooperative models of stereopsis rely on two fundamental rules first proposed by Marr and Poggio (1976). The *uniqueness* rule states that "each item from each image may be assigned at most one disparity value" and the *continuity* rule states that "disparity varies smoothly almost everywhere" as a consequence of the cohesiveness of matter, except at the boundaries of objects. A recent physiological study demonstrated the existence of local competitive and distant cooperative interactions in the primary visual cortex of the macaque, via lateral connections (Samonds, Potetz, & Lee, 2009). These interactions improve disparity sensitivity of binocular neurons over time. These authors suggest that local competition could be the neural substrate of the uniqueness rule, while distant cooperation would favor the detection of similar disparities and therefore implement the continuity rule. These horizontal connections should favor the detection of similar disparities in adjacent positions of the visual field and thus support the processing of 2D motion. It is possible that, in our stimuli, the lack of sensitivity to lateral motion is due to the implementation of the continuity rule. Because it is based on distant lateral connections, it can be hypothesized that this computation is slow.

The discrepancy between performances for 2D and 3D motion for our DO stimuli also has interesting implications in terms of predictive coding. It has been hypothesized that to reduce redundancy, the brain transmits only the unpredicted portions of the sensory input. This information is then combined with a predictive signal, boosting compatible inputs and discarding unlikely ones to reduce detection thresholds (Huang & Rao, 2011; Rao & Ballard, 1999; Srinivasan, Laughlin, & Dubs, 1982). Predictive coding has proven an adequate description of certain aspects of motion perception. For example, Roach, McGraw, and Johnston (2011) showed that a motion signal induces a prediction about the aspect and position of a forward stimulus and that this prediction is combined with the future representation of this stimulus. Our results suggest that the visual system might be more efficient in predicting the variations in depth than in lateral position of an object when it is defined only by binocular disparity.

Conclusion

In the present study, we measured optimal latencies for the perception of synchrony between moving visual stimuli and amplitude modulating sounds. We found that binocular vision is able to efficiently track variations in the direction of motion when these changes are variations in disparity/depth. However, we were surprised to find that this same system dedicated to process binocular vision seems to be poorly suited to track frontoparallel 2D motion. By using visual objects defined only by their binocular disparity, we were able to control for the level of processing required to compute audio-visual integration. Because disparity information is not available before early visual cortical areas, the optimal latencies measured in this study cannot result from early multimodal feedforward integration at a subcortical level.

References

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial vision*, *10*(4), 433–436.

Brooks, K. R. (2002). Interocular velocity difference contributes to stereomotion speed perception.

Journal of Vision, *2*(3):2, 218–231, http://www.journalofvision.org/content/2/3/2, doi:10.1167/2.3.2.

Brooks, K. R., & Stone, L. S. (2004). Stereomotion speed perception: Contributions from both changing disparity and interocular velocity difference over a range of relative disparities. *Journal of Vision, 4*(12):6, 1061–1079, http://www.journalofvision.org/content/4/12/6, doi:10.1167/4/12/6.

Cumming, B. G., & Parker, A. J. (1994). Binocular mechanisms for detecting motion-in-depth. *Vision Research*, *34*(4), 483–495.

Czuba, T. B., Rokers, B., Huk, A. C., & Cormack, L. K. (2010). Speed and eccentricity tuning reveal a central role for the velocity-based cue to 3D visual motion. *Journal of Neurophysiology*, *104*(5), 2886–2899.

Edwards, M., & Greenwood, J. A. (2005). The perception of motion transparency: A signal-to-noise limit. *Vision Research*, *45*(14), 1877–1884. doi:10.1016/j.visres.2005.01.026.

Greenwood, J. A., & Edwards, M. (2006). Pushing the limits of transparent-motion detection with binocular disparity. *Vision Research*, *46*(16), 2615–2624.

Harris, J. M., McKee, S. P., & Watamaniuk, S. N. (1998). Visual search for motion-in-depth: Stereomotion does not "pop out" from disparity noise. *Nature Neuroscience*, *1*(2), 165–168.

Harris, J. M., Nefs, H. T., & Grafton, C. E. (2008). Binocular vision and motion-in-depth. *Spatial Vision*, *21*(6), 531–547.

Harris, J. M., & Sumnall, J. H. (2000). Detecting binocular 3D motion in static 3D noise: No effect of viewing distance. *Spatial Vision*, *14*(1), 11–19.

Harris, J. M., & Watamaniuk, S. N. J. (1995). Speed discrimination of motion-in-depth using binocular cues. *Vision Research*, *35*(7), 885–896.

Hibbard, P. B., & Bradshaw, M. F. (1999). Does binocular disparity facilitate the detection of transparent motion? *Perception*, *28*(2), 183–191.

Huang, Y., & Rao, R. P. N. (2011). Predictive coding. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*(5), 580–593.

Julesz, B. (1960). Binocular depth perception of computer-generated patterns. *Bell System Technical Journal*, *39*, 1125–1162.

Lewald, J., & Guski, R. (2003). Cross-modal perceptual integration of spatially and temporally disparate auditory and visual stimuli. *Cognitive Brain Research*, *16*(3), 468–478.

Likova, L. T., & Tyler, C. W. (2007). Stereomotion processing in the human occipital cortex. *NeuroImage*, *38*(2), 293–305.

Mamassian, P., & Wallace, J. M. (2010). Sustained directional biases in motion transparency. *Journal of Vision*, *10*(13):23, 1–12, http://www.journalofvision.org/content/10/13/23, doi:10.1167/10.13.23.

Marr, D., & Poggio, T. (1976). Cooperative computation of stereo disparity. *Science*, *194*(4262), 283–287.

Maunsell, J. H., & Van Essen, D. C. (1983). Functional properties of neurons in middle temporal visual area of the macaque monkey. II. Binocular interactions and sensitivity to binocular disparity. *Journal of Neurophysiology*, *49*(5), 1148–1167.

Moutoussis, K., & Zeki, S. (1997a). A direct demonstration of perceptual asynchrony in vision. *Proceedings of the Royal Society of London B: Biological Sciences*, *264*(1380), 393–399.

Moutoussis, K., & Zeki, S. (1997b). Functional segregation and temporal hierarchy of the visual perceptive systems. *Proceedings of the Royal Society of London B: Biological Sciences*, *264*(1387), 1407–1414.

Nefs, H. T., & Harris, J. M. (2010). What visual information is used for stereoscopic depth displacement discrimination? *Perception*, *39*(6), 727–744. doi:10.1068/p6284.

Norcia, A. M., & Tyler, C. W. (1984). Temporal frequency limits for stereoscopic apparent motion processes. *Vision Research*, *24*(5), 395–401.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*(4), 437–442.

Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*(1), 79–87.

Rashbass, C., & Westheimer, G. (1961). Disjunctive eye movements. *Journal of Physiology*, *159*(2), 339–360.

Regan, D., & Beverley, K. (1973). Some dynamic features of depth perception. *Vision Research*, *13*(12), 2369–2379.

Roach, N. W., McGraw, P. V., & Johnston, A. (2011). Visual motion induces a forward prediction of spatial pattern. *Current Biology*, *21*(9), 740–745.

Rokers, B., Cormack, L. K., & Huk, A. C. (2009). Disparity- and velocity-based signals for three-dimensional motion perception in human MT+. *Nature Neuroscience*, *12*(8), 1050–1055.

Rokers, B., Czuba, T. B., Cormack, L. K., & Huk, A. C. (2011). Motion processing with two eyes in three dimensions. *Journal of Vision*, *11*(2):10, 1–19, http://www.journalofvision.org/content/11/2/10, doi:10.1167/11.2.10.

Samonds, J. M., Potetz, B. R., & Lee, T. S. (2009). Cooperative and competitive interactions facilitate stereo computations in macaque primary visual cortex. *Journal of Neuroscience*, *29*(50), 15780–15795.

Shioiri, S., Nakajima, T., Kakehi, D., & Yaguchi, H. (2008). Differences in temporal frequency tuning between the two binocular mechanisms for seeing motion in depth. *Journal of the Optical Society of America A: Optics & Image Science*, *25*(7), 1574–1585.

Snowden, R. J., & Rossiter, M. C. (1999). Stereoscopic depth cues can segment motion information. *Perception*, *28*(2), 193–201.

Srinivasan, M. V., Laughlin, S. B., & Dubs, A. (1982). Predictive coding: A fresh view of inhibition in the retina. *Proceedings of the Royal Society of London B: Biological Sciences*, *216*(1205), 427–459.

Stone, J. V., Hunkin, N. M., Porrill, J., Wood, R., Keeler, V., Beanland, M., Port, M., et al. (2001). When is now? Perception of simultaneity. *Proceedings of the Royal Society of London B: Biological Sciences]*, *268*(1462), 31–38.

Tyler, C. W. (1971). Stereoscopic depth movement: two eyes less sensitive than one. *Science*, *174*(4012), 958–961.

Uttal, W. R., Davis, N. S., & Welke, C. (1994). Stereoscopic perception with brief exposures. *Perception & Psychophysics*, *56*(5), 599–604.

Zeki, S., & Bartels, A. (1998). The asynchrony of consciousness. *Proceedings of the Royal Society of London B: Biological Sciences*, *265*(1405), 1583–1585.

VI The effect of audio-visual grouping on stereoacuity

The addition of auditory information in a visual task leads to significant facilitation in a various number of tasks such as visual search, motion perceptual learning and motion discrimination.

In the present study, we investigated whether grouping visual objects with a completely unrelated auditory signal (pitch variations) would affect sensitivity in the stereo domain. To do so, we measured stereoacuity (the smallest detectable depth difference that can be seen from binocular disparity) using lines distributed into two distinct depth planes. Lines from different depth planes could either be paired with a different pitch (congruent pairing condition) or with the same pitch (incongruent pairing condition). We manipulated the strength of the audio-visual grouping by varying the number of lines (one or three on each depth plane) in two separate experiments. Six participants were asked to focus on the two central lines of the display and to determine which line was nearer. They were instructed not to pay attention to the sound. Results showed a significant improvement (approximately 30%) of sensitivity in the congruent pairing condition compared to the incongruent pairing condition and to a control condition in which no sound was presented. We found no decrease in sensitivity in the incongruent pairing condition compared to the silent condition. Grouping in our stimuli led to substantial benefits but did not produce any cost. Our results suggest that a difference in pitch can improve stereoacuity, independent of the frequency content of the sound.

Key words: stereopsis, stereoacuity, multisensory integration, audio-visual facilitation.

1 Introduction

To achieve an optimal representation of a scene, the brain can make use of multiple sources of sensory information. Integrating from several sensory sources provides various advantages. For example, different senses provide complementary information (Burr & Alais, 2006; Ernst & Bülthoff, 2004). Combining redundant information from multiple sources is also an efficient way to reduce internal variability and increase the reliability of perceptual decisions (Ernst & Banks, 2002).

Before describing our own experiments, we will briefly review the literature on multisensory integration, focusing on examples from studies on audio-visual interactions.

1.1 Neurophysiology of multisensory integration

Evidence of multisensory integration at a subcortical level was primarily found in the superior colliculus (SC). This structure plays a role in orienting behaviours in response to covert and overt attention and receives ascending visual, auditory and somatosensory inputs. Neurons in the deep layers of the SC are often multimodal. Because the intrinsic role of the SC is to guide eye movements in response to various types of sensory stimulation, Meredith & Stein (1990) hypothesized the existence of multisensory integration mechanisms in this anatomical structure. They recorded the activity of such neurons and reported that when driven by spatially congruent stimuli they exhibit non-linear responses (Fig. VI.1), providing the first objective measure of multisensory integration. The amplitude of the multimodal response exceeds the sum of the unisensory components. These authors dubbed this effect *superadditivity* (Meredith & Stein, 2003). They also observed that superadditivity followed an *inverse effectiveness* rule: it is more likely to be observed when the unimodal inputs are weak. This principle of inverse effectiveness ensures the detection of weak stimulation and hence accurate and sensitive allocation of attention and eye movements.

Superadditivity in the SC therefore appears to be one of the earliest stages of multisensory optimization.

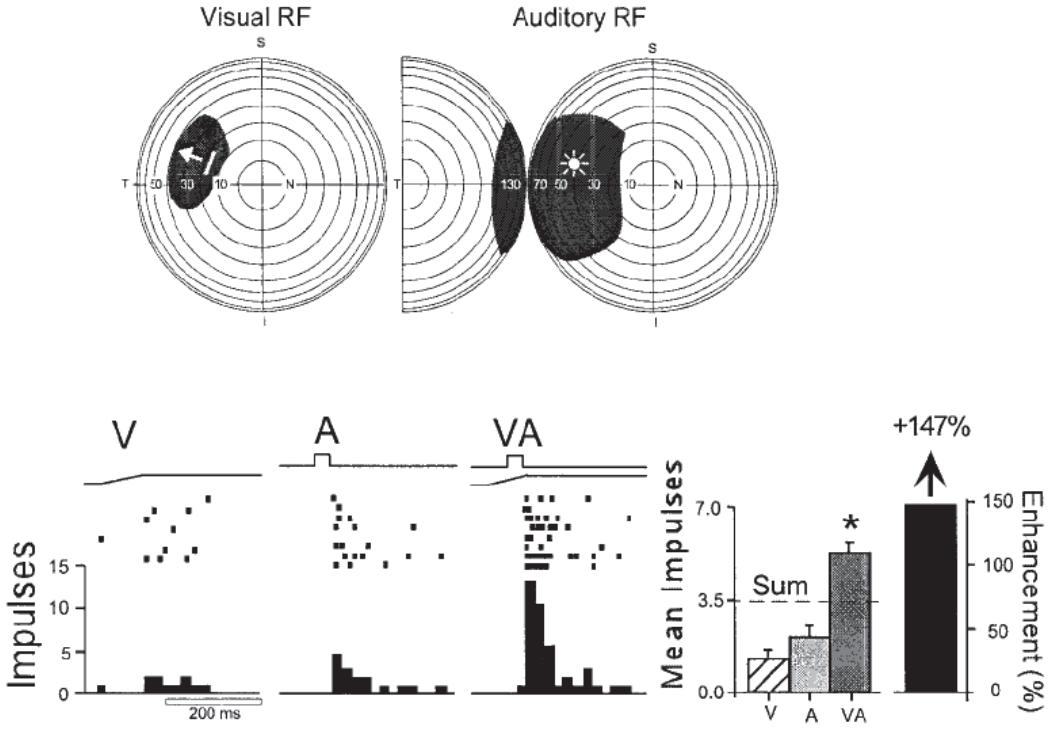


Figure VI.1 | Spatially coincident stimuli give rise to response enhancement. The top panels show the individual receptive fields (RFs) of this visual-auditory neuron from cat SC as gray-shaded areas on the diagrams of visual-auditory space. The position of each modality-specific stimulus is shown by an icon within the RF. The visual stimulus (V) was a moving bar of light whose direction of movement is indicated by the arrow. The auditory stimulus (A) was a broadband noise burst delivered from a stationary speaker. The bottom panels contain rasters and histograms illustrating the neuron's response to the modality-specific (visual alone, auditory alone) and multisensory (visual and auditory combined) stimuli, as well as bar graphs summarizing the mean responses and the index of multisensory enhancement. The spatially coincident visual-auditory pairing of stimuli resulted in a 147% response enhancement, well above the best modality-specific response and above the arithmetic sum of the two modality-specific responses (dashed line, t-test, $p < 0.05$). (reproduced from Calvert, Stein, & Spence, 2004)

At the cortical level, the traditional view that primary sensory cortices are sensory specific and functionally independent has been challenged by a number of studies

conducted in the last two decades (Driver & Noesselt, 2008; Fu et al., 2003; Schroeder & Foxe, 2002). One of the first demonstrations of cross-modal interactions in the cortex was provided by Calvert and colleagues (1997) who reported activation of auditory cortex during lip reading. The idea that sensory cortices are directly connected was backed up by a corpus of anatomical investigations and imaging studies on sensory deprivation. Anatomical investigations have revealed direct connections between the primary visual and auditory cortices (Cappe & Barone, 2005; Falchier, Clavagnier, Barone, & Kennedy, 2002). More specifically, it has been shown that auditory inputs in the primary visual cortex are distributed in the peripheral visual field. One possible advantage of this retinotopic distribution is the enhancement of spatial resolution, known to decrease with eccentricity from the foveal regions. It has been shown that primary visual cortex in blind individuals is activated during auditory, tactile and verbal tasks (Amedi, Raz, Pianka, Malach, & Zohary, 2003; Goyal, Hansen, & Blakemore, 2006; Kujala et al., 1995; Sadato et al., 1996) and that auditory cortex in deaf individuals is activated during visual tasks (Finney, Fine, & Dobkins, 2001). While spatio-temporal synchronization is necessary for superadditivity in the SC, multisensory integration in the cortex also seems to require congruence between the different sensory signals (Hein et al., 2007). Combining congruent multimodal signals might play a role in the identification of sensory stimulations into meaningful percepts (Andersen & Mamassian, 2008). Superadditivity has been found in superior temporal areas such as the left superior temporal gyrus (Foxe et al., 2002) and the left superior temporal sulcus (STS) (Calvert, Campbell, & Brammer, 2000). By varying signal strength, Stevenson & James (2009) demonstrated inverse effectiveness in the STS, suggesting strong superadditivity.

It is worth mentioning that most neuroimaging studies of higher cortical areas report small but reliable modulations of multisensory BOLD response that are not strong enough to qualify as superadditivity. For example, audio-visual and audio-tactile stimuli lead to an increase of BOLD response in STS of approximately 20% (Beauchamp, Lee, Argall, & Martin, 2004; Beauchamp, Yasar, Frye, & Ro, 2008; Newell, Mamassian, & Alais, 2010).

1.2 Behavioural measures of audio-visual integration

Because vision is traditionally considered as the dominant modality (Calvert et al., 2004), most studies on multisensory integration have focused on the effects of visual stimulation on other senses. More specifically, spatio-temporal integration of auditory and visual signals has been extensively investigated as a canonical example of multisensory integration. A key principle of multisensory integration is the modality appropriateness hypothesis: the modality that is most appropriate or reliable for a definite task dominates the perception in the context of that task. In the case of audio-visual integration, while audition displays greater temporal resolution and tends to dominate for duration judgment tasks, vision shows superior spatial resolution and dominates spatial localization tasks. Such a pattern of dominance can be revealed by presenting spatially or temporally incongruent audio-visual signals. For example, the illusory flash effect is a canonical example of dominance of audition over vision for temporal discrimination tasks. When a single flash is presented together with multiple auditory beeps it is perceived as multiple flashes (Shams, Kamitani, & Shimojo, 2000; 2002). Interestingly, this temporal alteration of vision by sound appears to be asymmetrical with respect to the total number of events. When multiple flashes are paired with a single beep, the illusion disappears, consistent with the idea that auditory temporal resolution is more reliable. Similarly, Recanzone (2003) showed that temporal visual rate perception is influenced by audition.

Conversely, the ventriloquist effect (Howard and Templeton, 1966) is the best-known example of vision's dominance: displacing a synchronized visual stimulus away from its corresponding sound source will produce a "capture" of the auditory stimulus by the visual event. However, Alais & Burr (2004) used a ventriloquism situation to demonstrate that, under specific circumstances, audition can dominate in spatial localization tasks. When the reliability of the visual signal is reduced by blurring the image, the perceived location of the audio-visual source is biased toward the auditory source.

In some cases where the input from one modality is ambiguous, information from another modality can be used to disambiguate (or even completely alter) the percept. For example, in the McGurk effect (1976), speech discrimination is altered by vision:

the sound of /ba/ is perceived as /da/ when it is presented with an image of a lip movement representing /ga/. In the stream/bounce illusion, the trajectory of two visual objects is deviated by adding a brief sound. In this situation, two disks oscillate back and forth across a square area and cross at the centre. When their trajectories cross, they can be perceived as bouncing apart or streaming past each other. The addition of a brief abrupt sound at the moment of impact is sufficient to bias the interpretation towards the bouncing percept.

1.3 Benefits of cross-modal interactions

Another way of looking at multisensory integration is to define situations in which a unimodal task is facilitated by the addition of a signal from another modality.

For example, audition has been shown to facilitate visual search (leading to shorter search times). Synchrony between a non-spatialized amplitude-modulating sound and a visual target modulating in luminance or depth presented among asynchronous distractors can efficiently guide visual search (van der Burg, Cass, Olivers, Theeuwes, & Alais, 2010; Zannoli, Cass, Mamassian, & Alais, 2012). In such experiments, correlating the sound with one of the distractors led to longer search times, suggesting that this facilitation might be the result of audio-visual integration and not solely cross-modal attention.

In several perceptual learning studies, Shams and colleagues found that a moving sound can substantially improve visual perceptual learning for motion discrimination tasks (Seitz, Kim, & Shams, 2006). Moreover, they found that this improvement of visual sensitivity with learning was significantly better when auditory and visual motion were congruent (in the same direction) (Kim, Seitz, & Shams, 2008). Because both congruent and incongruent conditions contained audio-visual stimuli, this facilitation could not be due to attention. These authors concluded that their results could be explained by multisensory interactions.

In a recent study, Kim, Peters & Shams (2012) showed that concurrent auditory stimuli improve accuracy in a motion detection task even though the auditory signal does not provide any useful information for the visual task. As in the perceptual learning studies presented above, this performance enhancement occurred only when

sound and visual motion moved in the same direction. The authors also concluded that their results could be explained by multisensory interactions.

1.4 Maximum-Likelihood Estimation

Currently, the most popular model used to describe how different types of information can be combined optimally is the Maximum-Likelihood Estimation (MLE) model (Alais & Burr, 2004; Ernst & Banks, 2002). According to the MLE model, the final estimate is a weighted linear sum of two or more signals that are weighted by their reliability. The more reliable, the more weight. Unimodal estimates are represented by a Gaussian function: the estimate is represented by the mean and the reliability is represented by the inverse of the variance. The mean of the final estimate is closer to the most reliable unimodal distribution and its variance is always inferior to the variance of the most optimal unimodal estimate. The MLE model captures some key ideas of multisensory integration: modality appropriateness and benefit from integration.

Ernst & Banks (2002) proposed a model to explain integration of two (or more) modalities when the two sensory signals should represent a common physical object. The MLE model, in addition to fitting well to various experimental configurations, provides a conceptualization of multisensory integration. Various sources of information about a single object reduce perceptual uncertainty and increase the precision of guided actions.

Another way of looking at multisensory integration is to study the interaction between signals that do not share a common source. For example, Otto & Mamassian (2012) investigated parallel decision processing with audio-visual signals using the redundant signal effect. They showed that multisensory decisions are made by accumulating evidence for each signal separately and that consequently more sensory noise is produced.

1.5 Aim of the present study

In the past recent years, research on multisensory integration has focused on demonstrating that cross-modal interactions could happen at very early stages of cortical sensory processing. The effects of auditory stimulation on visual perception described in the above section are in line with this goal. In the studies by Shams and colleagues on visual motion perception (Kim et al., 2012; 2008; Seitz et al., 2006), even though the auditory information was not critical for the task, congruency between auditory and visual signals was required. Accuracy improvement and perceptual learning facilitation fit with the general MLE framework: when two distinct pieces of information are available, the combined estimation is more reliable.

In the present study, we investigated whether the type of facilitation effects observed by Kim et al. (2012) would hold if the auditory and visual signals were related only by temporal correlation and not by congruency. To do so, we induced audio-visual grouping using an auditory cue that was orthogonal to the visual stimulation. In our stimuli, perceptual grouping was obtained by pairing visual objects with different pitches. We measured stereoacuity (the smallest detectable depth difference that can be seen from binocular disparity) as a function of audio-visual grouping. Because binocular disparity and auditory pitch do not share any perceptual congruency, we were able to test the effect of a completely orthogonal crossmodal signal on stereoacuity.

2 Method

We measured stereoacuity using the method of constant stimuli. The visual stimuli consisted of vertical lines presented sequentially from left to right or vice versa. Each line presentation was accompanied by an auditory beep. We manipulated the strength of the audio-visual (relative disparity / pitch) grouping by varying the number of elements in each trial. In the "weak grouping" experiment, two visual objects were presented while six objects were presented in the "strong grouping" experiment. In the two experiments, the lines were distributed into two distinct depth planes. For the strong grouping experiment, the lines were distributed in staggered

rows. The pairing between the two depth planes and the two pitches could be congruent (each depth plane was paired with a different pitch) or incongruent (two consecutive lines presented at different depths were paired with the same pitch)

2.1 Participants

The first experiment involved five participants (four naïve and one author). The second experiment involved six participants of which two also participated in the first experiment (including one author). All participants had normal or corrected-to-normal vision and were recruited from the laboratory. All had experience in psychophysical observation and had normal stereo acuity and hearing.

2.2 Stimulus presentation

Visual stimuli were presented on a CRT monitor (ViewSonic 21", resolution of 1280 x 960, refresh rate of 85 Hz) using a modified Wheatstone stereoscope at a simulated distance of 57 cm. Each eye viewed one horizontal half of the CRT screen. A chin rest was used to stabilize the observer's head and to control the viewing distance. The display was the only source of light and the stereoscope was calibrated geometrically to account for each participant's interocular distance. The auditory stimuli were presented binaurally through headphones.

2.3 Stimuli

2.3.1 Visual stimuli

Visual stimuli were generated using the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). They consisted of black (mean luminance 5 cd/m^2) lines (0.03 x 1 deg.) separated from each other by 0.2 deg and presented on a uniform grey background (4.6 x 4.6 deg — mean luminance 40 cd/m^2) at different depths. The depth of the lines was manipulated by adding opposite horizontal disparities to the left and right eyes images. The lines were evenly distributed around the centre of the background and presented sequentially (from left to right or vice versa) for 200 ms

with an inter stimulus interval of 100 ms in the first experiment and for 150 ms with an inter stimulus interval of 50 ms in the second (see Fig. VI.2).

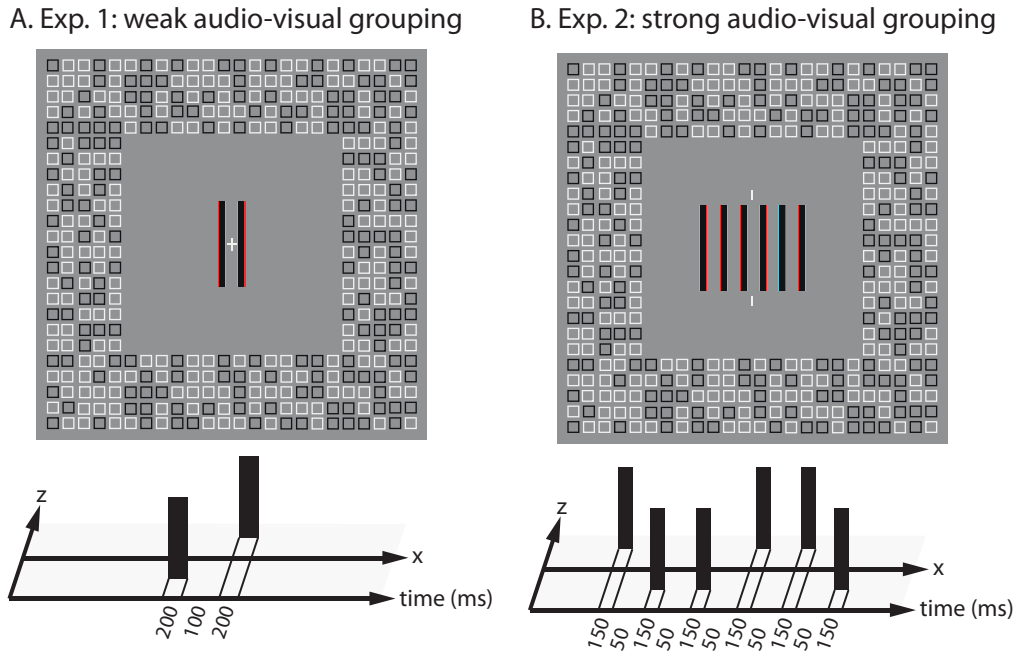


Figure VI.2 | Stimuli used in Experiment 1 (A.) and 2 (B.). The top row shows a binocular front view of the display, lines can be seen at different depths using anaglyph glasses. The bottom row represents the spatio-temporal configuration of the stimuli when the sequence starts to the left.

A vergence-stabilization frame was displayed on top of the background. It consisted of multiple luminance-defined squares (0.20 x 0.20 deg^2; black: 5 cd/m^2 and white: 80 cd/m^2). White nonius lines were presented at the centre of the display (see Figure).

2.3.2 Auditory stimuli

Auditory stimuli consisted of beeps of 400 Hz (low) and 600 Hz (high) with a duration of 200 ms in the first experiment and 150 ms in the second. Each line was presented together with a beep.

2.3.3 Disparity and pitch manipulations

In the second experiment, the six lines were systematically distributed in two depth planes as shown in Figure VI.3: the two central lines were always given opposite disparities. The depth difference between the two depth planes varied randomly between eight values. For the two experiments, four different experimental conditions were created by manipulating the association between relative disparity and pitch (see Fig. 4). Opposite disparities had the same probability to be associated either with the same pitch (low or high — 44.5% of the trials – "incongruent pairing" condition) or with a different pitch (44.5% of the trials – "congruent pairing" condition). In the "congruent pairing" condition, to avoid artificial perceptual learning of any type of association between disparity and pitch, the near plane could be associated either with the high or the low pitch and the far plane would be paired with the other pitch. These two sub-conditions were represented in the same proportions (22.2% of the trials for each condition). In the remaining 11.1% of the trials, no sound was presented. The four depth-pitch association conditions were interleaved. Each experiment contained a total of 864 trials and was divided into four blocks.

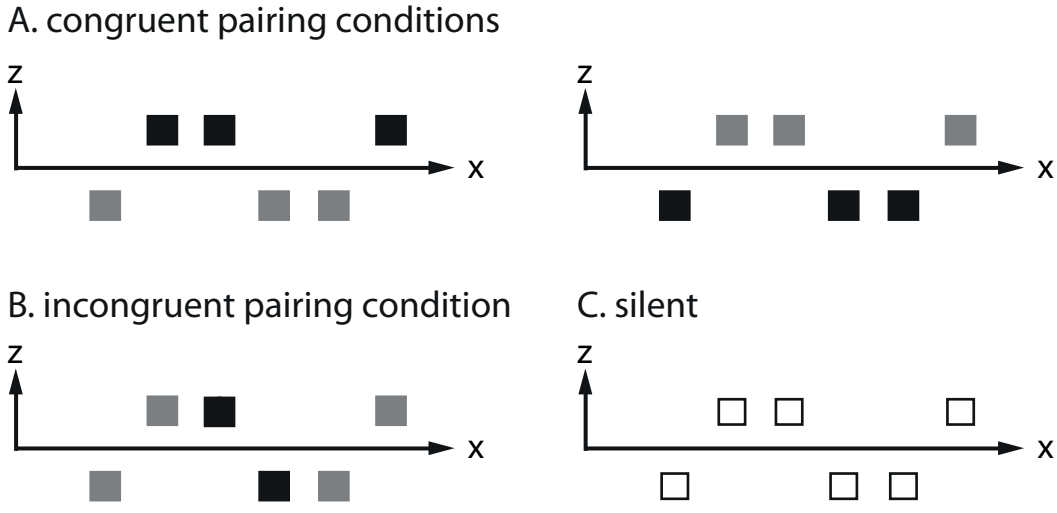


Figure VI.3 | Schematic representation of the association between disparity and pitch. The four panels represent a top view of the stimulus. The six lines (represented by squares) are displayed in Experiment 2 and only the two central lines are displayed in

Experiment 1. Light grey codes for the high pitch (660 Hz) and black codes for the low pitch (440 Hz).

Disparity values were chosen on the basis of preliminary pilot experiments to equate the subjective difficulty of the task across the four depth-pitch association conditions. A disparity pedestal, randomly chosen between +/- 2 arcmin, was added to the overall disparity of the lines. This manipulation ensured that the relative depth judgment task would rely on a comparison of the two depth planes and not on an absolute measure of the depth of only one depth plane compared to the plane if fixation.

2.4 Procedure

Each trial started with a presentation of the nonius lines (see Fig. VI.2). When correctly fusing the nonius, participants pressed any key to start the sequential presentation of the lines. The sequence went from left to right or vice versa and the direction was chosen randomly for each trial. Each trial lasted 300 ms for Experiment 1 and 1150 ms for Experiment 2. For Experiment 1, participants had to decide which of the two lines was in front of the other and respond using two different keys on a keyboard. In Experiment 2, participants had to focus on the two central lines and perform the same task as in Experiment 1. For experiment 1, nonius lines disappeared when the first stimulus line was presented while they stayed on for Experiment 2, to signal which lines were relevant for the relative depth judgment task.

3 Results

Psychometric functions were fitted to the proportion of right lines (relative to the nonius) seen in front as a function of the relative disparity between the two lines and thresholds were extracted for each sound condition (Palamedes toolbox). Figure VI.4 shows the thresholds as a function of the audio-visual pairing condition for Experiments 1 and 2.

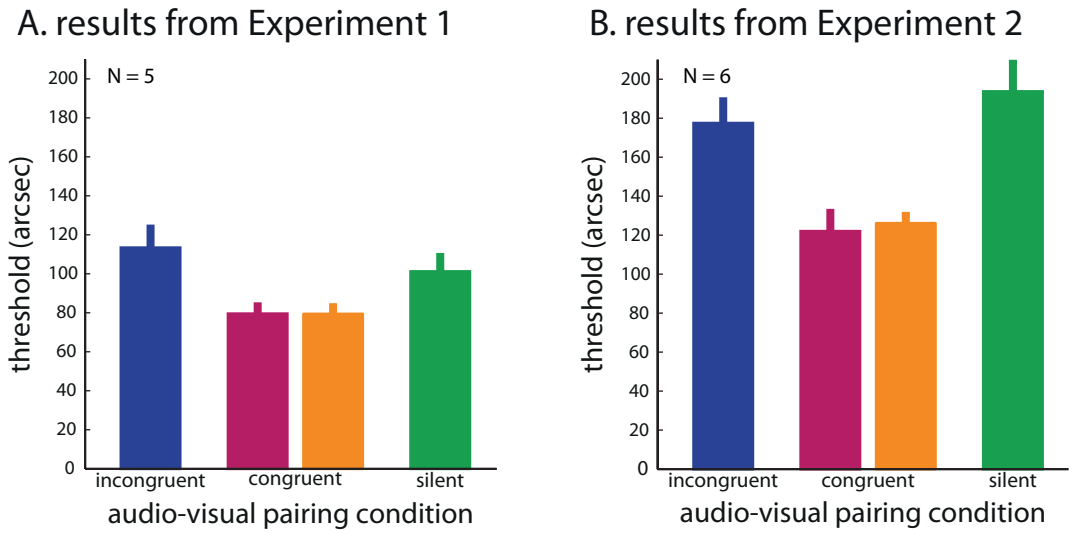


Figure VI.4 | Results from Experiments 1 and 2. Thresholds (arcsec) as a function of the audio-visual pairing. A. There are no significant differences between the four experimental conditions. B. There is no difference between the two congruent pairing conditions. Thresholds in these two conditions are significantly lower than in the incongruent pairing and silent conditions. The incongruent pairing and silent conditions are not significantly different.

One-way repeated-measures ANOVAs were run on the thresholds data for the two experiments. The ANOVA was not significant for Experiment 1 and significant for Experiment 2. To further investigate the effects in Experiment 2, we test multiple comparisons with Tukey least significant difference corrections. We found no difference between the silent and incongruent-pairing conditions and no difference between the two congruent-pairing conditions. All other comparisons were significant.

4 Discussion

In the two experiments reported here, we tested whether grouping visual objects with unrelated auditory information would affect stereo sensitivity. We found an increase in stereoacuity of approximately 30% when the two depth planes were segregated by pitch.

Pitch has been previously found to be a cue to depth for the localization of sound sources. Because of greater attenuation of high frequencies, a distant sound

carries more low frequencies. As the distance between the listener and a sound source increases, the sound is therefore perceived as having a lower pitch. In our stimuli, near depth was either paired with the low or the high pitch and vice versa for the far depth. Because these two different conditions were equally represented in the experiment, the design of our stimuli did not carry any artificial association between depth and pitch. Because we found no significant difference between our two congruent pairing conditions, we conclude that there was no cross-modal integration of disparity and pitch for the perception of depth based on stimulus congruency.

Experiments 1 and 2 show the same pattern of results. However, the difference between the congruent and incongruent / silent conditions is significant only in Experiment 2, suggesting that the strength of the perceptual grouping was a critical factor.

As described in the Introduction section, previous studies have investigated the effect of sound on various visual tasks. For example, Kim, Seitz & Shams (2008) examined the effect of auditory-visual congruency on visual learning. Participants were trained on a visual motion coherence detection task with either congruent (same direction) or incongruent (opposite direction) auditory stimuli and found that learning facilitation occurred only when auditory and visual motion signals were congruent. The authors concluded that this facilitation was subtended by multisensory integration. More recently, Kim, Peters & Shams (2012) developed a similar paradigm in which participants had to detect which of two intervals contained a coherent motion signal. They showed that adding an identical moving sound to both intervals improved accuracy but only when the auditory and visual motion signals were congruent. They concluded that this improvement in performance was due to audio-visual interactions at a sensory level. To our knowledge, an increase in visual sensitivity thanks to the addition of completely orthogonal non-informative auditory signal has never been reported.

We think that there are very low chances that such an increase in sensitivity is due to cross-modal attention processes. Because auditory and visual stimuli were presented simultaneously, it is unlikely that the pitch difference between the two sounds was used to anticipate a change in disparity.

This pattern of results relates to a series of observations made by Mamassian

(2008). In his study, pairs of vertical lines of same or opposite disparities were grouped by horizontal lines (creating slanted or flat rectangles) or by different contrasts. Discrimination thresholds were at least 10 times higher for lines belonging to the same group (same contrast or same rectangle), even though the disparity information was identical. In our study, lines were grouped by pitch: when they were associated with a different pitch, their relative disparity was easier to see. Our results are in line with Mamassian (2008), using information from a different modality (audition) to induce grouping.

Mamassian's results could be interpreted in terms of averaging. Depth information within a group is averaged and then compared to the average depth in the other group. Such a mechanism would be advantageous when the same disparities are grouped together: depth is estimated over several samples and then averaged, providing a more accurate estimate of depth (leading to lower thresholds). When opposite disparities are grouped together, the average disparity is null: in this case averaging has detrimental effects on the discrimination task (leading to higher thresholds). However, we did not find any impairment in the incongruent pairing condition compared to the silent condition. This lack of significance might be the result of confounding effects in the incongruent pairing condition. The detrimental effect of grouping in this condition might have been rubbed out by a general reduction of temporal uncertainty in the audio-visual conditions compared to the silent condition. The auditory sequence of beeps could sharpen the perception of the visual onsets and offsets. This could have led to a significant increase in overall sensitivity in the audio-visual pairing conditions compared to the silent condition. To test this possibility, it might be interesting to run a control condition in which pitch values (either 440 Hz or 660 Hz) would be attributed randomly for each visual object. This way, no systematic grouping is induced but the auditory information can still be used to lower the temporal uncertainty of the visual events. If the grouping hypothesis holds, we expect thresholds in the control random condition to fall between the congruent and incongruent pairing conditions.

To further investigate the effect of grouping on stereoacuity it would be interesting to manipulate the strength of the grouping on a trial-by-trial basis. This could be done by varying the proportion of congruent- and incongruent-pairing

within a sequence.

5 Conclusion

In the present study, we tested whether grouping visual objects by pitch would affect sensitivity in the stereo domain. We measured stereoacuity using vertical lines distributed into two depth planes. When the audio-visual pairing was congruent with the two depth planes we expected an increase in sensitivity whereas we expected a decrease in sensitivity when the audio-visual pairing was incongruent. We partly confirmed this prediction by finding that thresholds in the congruent pairing conditions were significantly smaller (of approximately 30%) than in the incongruent and silent conditions. This result demonstrates that the facilitation observed here is independent of the information content of the auditory signal suggesting that the mere presence of a pitch difference is sufficient for facilitation.

VII General discussion and conclusion

In the present thesis, we presented four distinct experimental projects that all aimed at understanding how the processing of binocular disparity can be affected by different types of non-stereoscopic information.

In a first series of psychophysical studies, we investigated how monocular regions are treated by the stereoscopic system and integrated with binocular disparity information to build the disparity map. To do so, we tested whether da Vinci stereopsis could be affected by the transparency of the occluding surface. We found that the position of monocular objects in depth was not sensitive to the material properties of objects, suggesting that da Vinci stereopsis is solved at relatively early stages of binocular disparity processing. Furthermore, a careful examination of the distribution of depth estimations across our experimental conditions suggested that the resolution of da Vinci stereopsis is underlined by a combination of classical stereoscopic mechanisms, occlusion constraints and a prior preference for small disparities. In other words, the spatial arrangement of monocular features in the image can be efficiently used by the visual system to refine the shape of the disparity map.

In a second series of experiments, we tested whether a non-spatial auditory signal could improve visual search in the disparity domain. For stimuli defined exclusively by stereomotion, we found that square-wave amplitude modulations correlated with the depth modulation of the target object could efficiently drive visual search. These results suggest that a temporally correlated sound signal can be used by stereomotion detectors to process the change of disparity over time.

In a third series of experiments, we investigated motion discrimination in the 2D and 3D domain. We measured the optimal latency for the perception of synchrony between an amplitude-modulating sound and visual stimuli moving laterally or in depth. We found that the optimal latency for the perception of synchrony for 3D motion was similar whether the stimuli were defined by luminance or disparity, suggesting that the processing of binocular disparity can be substantially

fast. Surprisingly, we found that the discrimination of lateral motion for stimuli defined exclusively by disparity was much worse than for motion-in-depth. These results suggest that stereomotion detectors are poorly suited to track 2D motion

In a fourth series of experiments, we investigated the influence of audio-visual grouping on stereoacuity. We found that a non-informative orthogonal sound signal presented concurrently with the disparity information could improve stereoacuity by approximately 30% when two depth planes were segregated by sound. We expected that averaging of disparity information according to audio-visual grouping would have produced impairment in a condition in which different depths were paired with identical pitches. We did not observe this detrimental effect in our data and we suspect that it might have been rubbed out by a general reduction of temporal uncertainty in vision using the auditory signal in the two audio-visual conditions. Further testing is required to confirm this hypothesis. The design and results in the different experimental conditions of the experiments allowed us to discard the potential role of cross-modal information.

Taken together, the results exposed in this thesis strongly support the general idea that the stereoscopic system is not fully encapsulated and works in cooperation with other within-vision and auditory processes to increase its spatial and temporal precision.

References

Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current biology*, *14*(3), 257–262.

Alais, D., Cass, J., O'Shea, R. P., & Blake, R. (2010). Visual sensitivity underlying changes in visual consciousness. *Current biology*, *20*(15), 1362–1367.

Allison, R. S., Howard, I. P., & Howard, A. (1998). Motion in depth can be elicited by dichoptically uncorrelated textures. *Perception*, *2* (S), 46.

Almeida, J., Mahon, B. Z., Nakayama, K., & Caramazza, A. (2008). Unconscious processing dissociates along categorical lines. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(39), 15214–15218.

Amedi, A., Raz, N., Pianka, P., Malach, R., & Zohary, E. (2003). Early "visual" cortex activation correlates with superior verbal memory performance in the blind. *Nature neuroscience*, *6*(7), 758–766.

Andersen, T. S., & Mamassian, P. (2008). Audiovisual integration of stimulus transients. *Vision Research*, *48*(25), 2537–2544.

Anderson, B. L. (1994). The role of partial occlusion in stereopsis. *Nature*, *367*(6461), 365–368.

Anderson, B. L., & Nakayama, K. (1994). Toward a general theory of stereopsis: binocular matching, occluding contours, and fusion. *Psychological Review*, *101*(3), 414–445.

Anderson, P. A., & Movshon, J. A. (1989). Binocular combination of contrast signals. *Vision Research*, *29*(9), 1115–1132.

Andrews, T. J., Glennerster, A., & Parker, A. J. (2001). Stereoacuity thresholds in the presence of a reference surface. *Vision Research*, *41*(23), 3051–3061.

Assee, A., & Qian, N. (2007). Solving da Vinci stereopsis with depth-edge-selective V2 cells. *Vision Research*, *47*(20), 2585–2602.

Backus, B. T., & Banks, M. S. (1999). Estimator reliability and distance scaling in stereoscopic slant perception. *Perception*, *28*(2), 217–242.

Backus, B. T., Banks, M. S., van Ee, R., & Crowell, J. A. (1999). Horizontal and vertical disparity, eye position, and stereoscopic slant perception. *Vision Research*, *39*(6), 1143–1170.

Backus, B. T., Fleet, D. J., Parker, A. J., & Heeger, D. J. (2001). Human cortical activity correlates with stereoscopic depth perception. *Journal of Neurophysiology*, *86*(4), 2054–2068.

Ban, H., Preston, T. J., Meeson, A., & Welchman, A. E. (2012). The integration of motion and disparity cues to depth in dorsal visual cortex. *Nature neuroscience*, *15*(4), 636–643.

Banks, M. S., Gepshtein, S., & Landy, M. S. (2004). Why is spatial stereoresolution so low? *Journal of Neuroscience*, *24*(9), 2077–2089.

Barlow, H. B., Blakemore, C., & Pettigrew, J. D. (1967). The neural mechanism of binocular depth discrimination. *The Journal of physiology*, *193*(2), 327–342.

Beauchamp, M. S., Lee, K. E., Argall, B. D., & Martin, A. (2004). Integration of auditory and visual information about objects in superior temporal sulcus. *Neuron*, *41*(5), 809–823.

Beauchamp, M. S., Yasar, N. E., Frye, R. E., & Ro, T. (2008). Touch, sound and

vision in human superior temporal sulcus. *NeuroImage*, *41*(3), 1011–1020.
Beverley, K. I., & Regan, D. (1973). Evidence for the existence of neural mechanisms selectively sensitive to the direction of movement in space. *The Journal of physiology*, *235*(1), 17–29.
Blake, R. (1989). A neural theory of binocular rivalry. *Psychological Review*, *96*(1), 145–167.
Blake, R., & Fox, R. (1973). The psychophysical inquiry into binocular summation. *Attention, perception & psychophysics*, *14*(1), 161–185.
Blake, R., Sobel, K. V., & Gilroy, L. A. (2003). Visual motion retards alternations between conflicting perceptual interpretations. *Neuron*, *39*(5), 869–878.
Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial vision*, *10*(4), 433–436.
Bredfeldt, C. E., & Cumming, B. G. (2006). A simple account of cyclopean edge responses in macaque v2. *Journal of Neuroscience*, *26*(29), 7581–7596.
Brooks, K. (2001). Stereomotion speed perception is contrast dependent. *Perception*, *30*(6), 725–731.
Brooks, K. R. (2002a). Interocular velocity difference contributes to stereomotion speed perception. *Journal of Vision*, *2*(3), 218–231.
Brooks, K. R. (2002b). Monocular motion adaptation affects the perceived trajectory of stereomotion. *Journal of Experimental Psychology: Human Perception and Performance*, *28*(6), 1470–1482.
Brooks, K. R., & Stone, L. S. (2004). Stereomotion speed perception: Contributions from both changing disparity and interocular velocity difference over a range of relative disparities. *Journal of Vision*, *4*(12), 6–6.
Brooks, K. R., & Stone, L. S. (2006). Spatial scale of stereomotion speed processing. *Journal of Vision*, *6*(11), 9–9.
Brooks, K., & Gillam, B. (2006). Quantitative perceived depth from sequential monocular decamouflage. *Vision Research*, *46*(5), 605–613.
Burr, D., & Alais, D. (2006). Combining visual and auditory information (Vol. 155, pp. 243–258). Presented at the Visual Perception, Pt 2: Fundamentals of Awareness: Multi-Sensory Integration and High-Order Perception.
Burt, P., & Julesz, B. (1980). A disparity gradient limit for binocular fusion. *Science*, *208*(4444), 615–617.
Calvert, G. A., Bullmore, E. T., Brammer, M. J., Campbell, R., Williams, S. C. R., McGuire, P. K., Woodruff, P. W. R., et al. (1997). Activation of auditory cortex during silent lipreading. *Science*, *276*(5312), 593–596.
Calvert, G. A., Campbell, R., & Brammer, M. J. (2000). Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Current Biology*, *10*(11), 649–657.
Calvert, G. A., Stein, B. E., & Spence, C. (2004). The handbook of multisensory processes. MIT Press.
Campbell, F. W., & Green, D. G. (1965). Monocular versus Binocular Visual Acuity. *Nature*, *208*(5006), 191–192.
Cappe, C., & Barone, P. (2005). Heteromodal connections supporting multisensory integration at low levels of cortical processing in the monkey. *European Journal of Neuroscience*, *22*(11), 2886–2902.
Carlson, T. A., & He, S. (2000). Visible binocular beats from invisible monocular stimuli during binocular rivalry. *Current biology* , *10*(17), 1055–1058.
Chandrasekaran, C., Canon, V., Dahmen, J. C., Kourtzi, Z., & Welchman, A. E.

(2007). Neural correlates of disparity-defined shape discrimination in the human brain. *Journal of Neurophysiology*, *97*(2), 1553–1565.
Chong, S. C., Tadin, D., & Blake, R. (2005). Endogenous attention prolongs dominance durations in binocular rivalry. *Journal of Vision*, *5*(11), 1004–1012.
Chopin, A., & Mamassian, P. (2012). Predictive properties of visual adaptation. *Current biology*, *22*(7), 622–626.
Cook, M., & Gillam, B. (2004). Depth of Monocular Elements in a Binocular Scene: The Conditions for da Vinci Stereopsis. *Journal of Experimental Psychology: Human Perception and Performance*, *30*(1), 92–103.
Cormack, L. K., Stevenson, S. B., & Schor, C. M. (1991). Interocular correlation, luminance contrast and cyclopean processing. *Vision Research*, *31*(12), 2195–2207.
Cumming, B. (1995). The relationship between stereoacuity and stereomotion thresholds. *Perception*, *24*(1), 105–114.
Cumming, B. (1999). Binocular Neurons in V1 of Awake Monkeys Are Selective for Absolute, Not Relative, Disparity. *The Journal of neuroscience*.
Cumming, B. G., & DeAngelis, G. C. (2001). The physiology of stereopsis. *Annual review of neuroscience*, *24*, 203–238.
Cumming, B. G., & Parker, A. J. (1994). Binocular mechanisms for detecting motion-in-depth. *Vision Research*, *34*(4), 483–495.
Cumming, B. G., & Parker, A. J. (1997). Responses of primary visual cortical neurons to binocular disparity without depth perception. *Nature*, *389*(6648), 280–283.
Cynader, M., & Regan, D. (1978). Neurones in cat parastriate cortex sensitive to the direction of motion in three-dimensional space. *The Journal of physiology*, *274*, 549–569.
Cynader, M., & Regan, D. (1982). Neurons in cat visual cortex tuned to the direction of motion in depth: effect of positional disparity. *Vision Research*, *22*(8), 967–982.
Czuba, T. B., Rokers, B., Guillet, K., Huk, A. C., & Cormack, L. K. (2011). Three-dimensional motion aftereffects reveal distinct direction-selective mechanisms for binocular processing of motion through depth. *Journal of Vision*, *11*(10), 18–18.
Czuba, T. B., Rokers, B., Huk, A. C., & Cormack, L. K. (2010). Speed and Eccentricity Tuning Reveal a Central Role for the Velocity-Based Cue to 3D Visual Motion. *Journal of Neurophysiology*, *104*(5), 2886–2899.
der Burg, Van, E., Cass, J., Olivers, C. N. L., Theeuwes, J., & Alais, D. (2010). Efficient Visual Search from Synchronized Auditory Signals Requires Transient Audiovisual Events. *PLoS ONE*, *5*(5), e10664.
Diaz-Caneja, E. (1928). Sur l'alternance binoculaire. *Annales D'Oculistique*, *165*(October), 721–731.
Ding, J., & Levi, D. M. (2011). Recovery of stereopsis through perceptual learning in human adults with abnormal binocular vision. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(37), E733–41.
Ding, J., & Sperling, G. (2006). A gain-control theory of binocular combination. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(4), 1141–1146.
Driver, J., & Noesselt, T. (2008). Multisensory Interplay Reveals Crossmodal Influences on "Sensory-Specific" Brain Regions, Neural Responses, and Judgments. *Neuron*, *57*(1), 11–23.
Durand, J.-B., Celebrini, S., & Trotter, Y. (2007). Neural bases of stereopsis across

visual field of the alert macaque monkey. *Cerebral cortex (New York, N.Y. : 1991)*, *17*(6), 1260–1273.
Durand, J.-B., Zhu, S., Celebrini, S., & Trotter, Y. (2002). Neurons in parafoveal areas V1 and V2 encode vertical and horizontal disparities. *Journal of Neurophysiology*, *88*(5), 2874–2879.
Erkelens, C. J., & Collewijn, H. (1985). Motion perception during dichoptic viewing of moving random-dot stereograms. *Vision Research*, *25*(4), 583–588.
Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*(6870), 429–433.
Ernst, M. O., & Bülthoff, H. H. (2004). Merging the senses into a robust percept. *Trends in cognitive sciences*, *8*(4), 162–169.
Falchier, A., Clavagnier, S., Barone, P., & Kennedy, H. (2002). Anatomical evidence of multimodal integration in primate striate cortex. *Journal of Neuroscience*, *22*(13), 5749–5759.
Fernandez, J. M., & Farell, B. (2005). Seeing motion in depth using inter-ocular velocity differences. *Vision Research*, *45*(21), 2786–2798.
Filippini, H. R., & Banks, M. S. (2009). Limits of stereopsis explained by local cross-correlation. *Journal of Vision*, *9*(1), 8.1–18.
Finney, E. M., Fine, I., & Dobkins, K. R. (2001). Visual stimuli activate auditory cortex in the deaf. *Nature neuroscience*, *4*(12), 1171–1173.
Foxe, J. J., Wylie, G. R., Martinez, A., Schroeder, C. E., Javitt, D. C., Guilfoyle, D., Ritter, W., et al. (2002). Auditory-somatosensory multisensory processing in auditory association cortex: an fMRI study. *Journal of Neurophysiology*, *88*(1), 540–543.
Freeman, A. W. (2005). Multistage model for binocular rivalry. *Journal of Neurophysiology*, *94*(6), 4412–4420.
Fu, K.-M. G., Johnston, T. A., Shah, A. S., Arnold, L., Smiley, J., Hackett, T. A., Garraghty, P. E., et al. (2003). Auditory cortical neurons respond to somatosensory stimulation. *Journal of Neuroscience*, *23*(20), 7510–7515.
Geiger, D., & Ladendorf, B. (1995). Occlusions and binocular stereo. *International Journal of Computer Vision*.
Gillam, B., & Lawergren, B. (1983). The induced effect, vertical disparity, and stereoscopic theory. *Perception & psychophysics*, *34*(2), 121–130.
Gillam, B., Blackburn, S., & Cook, M. (1995). Panum's limiting case: double fusion, convergence error, or 'da Vinci stereopsis'. *Perception*, *24*(3), 333–346.
Gillam, B., Blackburn, S., & Nakayama, K. (1999). Stereopsis based on monocular gaps: metrical encoding of depth and slant without matching contours. *Vision Research*, *39*(3), 493–502.
Gillam, B., & Borsting, E. (1988). The role of monocular regions in stereoscopic displays. *Perception*, *17*(5), 603–608.
Gillam, B., Cook, M., & Blackburn, S. (2003). Monocular discs in the occlusion zones of binocular surfaces do not have quantitative depth -- a comparison with Panum's limiting case. *Perception*, *32*(8), 1009–1019.
Gonzalez, F., Justo, M. S., Bermudez, M. A., & Perez, R. (2003). Sensitivity to horizontal and vertical disparity and orientation preference in areas V1 and V2 of the monkey. *Neuroreport*, *14*(6), 829–832.
Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences*, *15*(1), 20–25.

Goutcher, R., & Mamassian, P. (2005). Selective biasing of stereo correspondence in an ambiguous stereogram. *Vision Research*, *45*(4), 469–483.
Goyal, M. S., Hansen, P. J., & Blakemore, C. B. (2006). Tactile perception recruits functionally related visual areas in the late-blind. *Neuroreport*, *17*(13), 1381–1384.
Grossberg, S., & Howe, P. D. L. (2003). A laminar cortical model of stereopsis and three-dimensional surface perception. *Vision Research*, *43*(7), 801–829.
Grove, P. M., Ben Sachtler, W. L., & Gillam, B. J. (2006). Amodal completion with background determines depth from monocular gap stereopsis. *Vision Research*, *46*(22), 3771–3774.
Grove, P. M., Gillam, B., & Ono, H. (2002). Content and context of monocular regions determine perceived depth in random dot, unpaired background and phantom stereograms. *Vision Research*, *42*(15), 1859–1870.
Harris, J. M., & Sumnall, J. H. (2000). Detecting binocular 3D motion in static 3D noise: no effect of viewing distance. *Spatial vision*, *14*(1), 11–19.
Harris, J. M., & Watamaniuk, S. N. J. (1995). Speed discrimination of motion-in-depth using binocular cues. *Vision Research*, *35*(7), 885–896.
Harris, J. M., McKee, S. P., & Smallman, H. S. (1997). Fine-scale processing in human binocular stereopsis. *Journal of the Optical Society of America A, Optics, image science, and vision*, *14*(8), 1673–1683.
Harris, J. M., McKee, S. P., & Watamaniuk, S. N. (1998). Visual search for motion-in-depth: stereomotion does not "pop out" from disparity noise. *Nature neuroscience*, *1*(2), 165–168.
Harris, J. M., Nefs, H. T., & Grafton, C. E. (2008). Binocular vision and motion-in-depth. *Spatial vision*, *21*(6), 531–547.
Hayashi, R., Maeda, T., Shimojo, S., & Tachi, S. (2004). An integrative model of binocular vision: a stereo model utilizing interocularly unpaired points produces both depth and binocular rivalry. *Vision Research*, *44*(20), 2367–2380.
Häkkinen, J., & Nyman, G. (1996). Depth asymmetry in da Vinci stereopsis. *Vision Research*, *36*(23), 3815–3819.
He, Z. J., & Nakayama, K. (1995). Visual attention to surfaces in three-dimensional space. *Proceedings of the National Academy of Sciences of the United States of America*, *92*(24), 11155–11159.
Hein, G., Doehrmann, O., Müller, N. G., Kaiser, J., Muckli, L., & Naumer, M. J. (2007). Object familiarity and semantic congruency modulate responses in cortical audiovisual integration areas. *Journal of Neuroscience*, *27*(30), 7881–7887.
Hering, E. (1861). Beiträge zur Physiologie. *Leipzig: Engelmann*.
Howard, I. P. (2002). Seeing in depth (Vol. I). Porteous Press.
Howard, I. P., & Rogers, B. J. (2002). Seeing in depth (Vol. II). Porteous Press.
Hubel, D. H., & Wiesel, T. N. (1959). Receptive fields of single neurones in the cat's striate cortex. *The Journal of physiology*, *148*, 574–591.
Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of physiology*, *160*, 106–154.
Humphriss, D. (1982). The psychological septum – An investigation into its function. *American Journal Of Optometry And Physiological Optics (59)*, 639-41.
Janssen, P., Vogels, R., & Orban, G. A. (2000). Three-dimensional shape coding in inferior temporal cortex. *Neuron*, *27*(2), 385–397.
Janssen, P., Vogels, R., Liu, Y., & Orban, G. A. (2003). At least at the level of

inferior temporal cortex, the stereo correspondence problem is solved. *Neuron*, *37*(4), 693–701.
Jiang, Y., & He, S. (2006). Cortical Responses to Invisible Faces: Dissociating Subsystems for Facial-Information Processing. *Current Biology*, *16*(20), 2023–2029.
Jiang, Y., Costello, P., & He, S. (2007). Processing of invisible stimuli: advantage of upright faces and recognizable words in overcoming interocular suppression. *Psychological science*, *18*(4),
Jones, D. G., & Malik, J. (1992). Computational framework for determining stereo correspondence from a set of linear spatial filters. *Image and Vision Computing*, *10*(10), 699–708.
Jones, R. K., & Lee, D. N. (1981). Why two eyes are better than one: the two views of binocular vision. *Journal of Experimental Psychology: Human Perception and Performance*, *7*(1), 30–40.
Julesz, B. (1960). Binocular depth perception of computer-generated patterns. *Bell System Technical Journal, 39*, 1125-1162.
Julesz, B. (1964a). Binocular depth perception without familiarity cues. *Science*, *145*, 356–362.
Julesz, B. (1964b). Binocular depth perception without familiarity cues. *Science*, *145*(3630), 356.
Julesz, B., & Tyler, C. W. (1976). Neurontropy, an entropy-like measure of neural correlation, in binocular fusion and rivalry. *Biological cybernetics*, *23*(1), 25–32.
Kanade, T., & Okutomi, M. (1991). A stereo matching algorithm with an adaptive window: Theory and experiment. *Robotics and Automation, 1991. Proceedings., 1991 IEEE International Conference on*, 1088–1095 vol. 2.
Kaufman, L. (1964). Suppression and fusion in viewing complex stereograms. *The American Journal of Psychology*, *77*, 193–205.
Kim, R. S., Seitz, A. R., & Shams, L. (2008). Benefits of Stimulus Congruency for Multisensory Facilitation of Visual Learning. (M. Herzog, Ed.)*PLoS ONE*, *3*(1), e1532.
Kim, R., Peters, M. A. K., & Shams, L. (2012). 0 + 1 > 1: How Adding Noninformative Sound Improves Performance on a Visual Task. *Psychological Science*, *23*(1), 6–12.
Kitaoji, H., & Toyama, K. (1987). Preservation of position and motion stereopsis in strabismic subjects. *Investigative Ophthalmology & Visual Science*, *28*(8), 1260–1267.
Kovács, I., Papathomas, T. V., Yang, M., & Fehér, Á. (1996). When the brain changes its mind: interocular grouping during binocular rivalry. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(26), 15508–15511.
Kujala, T., Huotilainen, M., Sinkkonen, J., Ahonen, A. I., Alho, K., Hämäläinen, M. S., Ilmoniemi, R. J., et al. (1995). Visual cortex activation in blind humans during sound discrimination. *Neuroscience Letters*, *183*(1-2), 143–146.
Lawson, R. B., & Gulick, W. L. (1967). Stereopsis and anomalous contour. *Vision Research*, *7*(3), 271–297.
Legge, G. E. (1984a). Binocular contrast summation—I. Detection and discrimination. *Vision Research*, *24*(4), 373–383.
Legge, G. E. (1984b). Binocular contrast summation--II. Quadratic summation. *Vision Research*, *24*(4), 385–394.

Lehky, S. R. (1988). An astable multivibrator model of binocular rivalry. *Perception*, *17*(2), 215–228.
Levelt, W. (1966). The alternation process in binocular rivalry. *British Journal of Psychology*, *57*(3-4), 225–238.
Levelt, W. J. (1965). Binocular Brightness Averaging and Contour Information. *British journal of psychology*, *56*, 1–13.
Likova, L. T., & Tyler, C. W. (2007). Stereomotion processing in the human occipital cortex. *NeuroImage*, *38*(2), 293–305.
Maeda, M., Sato, M., Ohmura, T., Miyazaki, Y., Wang, A., & Awaya, S. (1999). Binocular depth-from-motion in infantile and late-onset esotropia patients with poor stereopsis. *Investigative Ophthalmology & Visual Science*, *40*(12), 3031–3036.
Mamassian, P. (2008). Depth, but not surface orientation, from binocular disparities. *Journal of Vision, 8* (6).
Marr, & Poggio. (1979). A Theory of Human Stereo Vision, 1–89.
Marr, D., & Poggio, T. (1976). Cooperative computation of stereo disparity. *Science*, *194*(4262), 283–287.
Masson, G. S., Busettini, C., & Miles, F. A. (1997). Vergence eye movements in response to binocular disparity without depth perception. *Nature*, *389*(6648), 283–286.
Maunsell, J. H., & Van Essen, D. C. (1983). Functional properties of neurons in middle temporal visual area of the macaque monkey. II. Binocular interactions and sensitivity to binocular disparity. *Journal of Neurophysiology*, *49*(5), 1148–1167.
Mayhew, J., & Longuet-Higgins, H. (1982). A computational model of binocular depth perception. *Nature*, *297*(5865), 376–378.
Meese, T. S., Georgeson, M. A., & Baker, D. H. (2006). Binocular contrast vision at and above threshold. *Journal of Vision*, *6*(11), 1224–1243.
Meredith, M. A., & Stein, B. E. (1990). The visuotopic component of the multisensory map in the deep laminae of the cat superior colliculus. *The Journal of neuroscience*, *10*(11), 3727–3742.
Meredith, M. A., & Stein, B. E. (2003). *The Merging of the Senses.* MIT Press, Cambridge, MA, USA.
Mueller, T. J., & Blake, R. (1989). A fresh look at the temporal dynamics of binocular rivalry. *Biological cybernetics*, *61*(3), 223–232.
Nakayama, K., & Shimojo, S. (1990). da Vinci stereopsis: depth and subjective occluding contours from unpaired image points. *Vision Research*, *30*(11), 1811–1825.
Neri, P., Bridge, H., & Heeger, D. (2004). Stereoscopic processing of absolute and relative disparity in human visual cortex. *Journal of Neurophysiology*, *92*(3), 1880–1891.
Newell, F., Mamassian, P., & Alais, D. (2010). Multisensory Processing in Review: from Physiology to Behaviour. *Seeing and perceiving*, *23*(1), 3–38.
Nienborg, H., & Cumming, B. G. (2006). Macaque V2 neurons, but not V1 neurons, show choice-related activity. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *26*(37), 9567–9578.
Nienborg, H., Bridge, H., Parker, A. J., & Cumming, B. G. (2004). Receptive field size in V1 neurons limits acuity for perceiving disparity modulation. *Journal of Neuroscience*, *24*(9), 2065–2076.

Norcia, A. M., & Tyler, C. W. (1984). Temporal frequency limits for stereoscopic apparent motion processes. *Vision Research*, *24*(5), 395–401.
Ogle, K. N. (1952). On the limits of stereoscopic vision. *Journal of experimental psychology*, *44*(4), 253–259.
Ogle, K. N., & Weil, M. P. (1958). Stereoscopic vision and the duration of the stimulus. *A.M.A. archives of ophthalmology*, *59*(1), 4–17.
Ohzawa, I., DeAngelis, G. C., & Freeman, R. D. (1997). Encoding of binocular disparity by complex cells in the cat's visual cortex. *Journal of Neurophysiology*, *77*(6), 2879–2909.
Ono, H., Lillakas, L., Grove, P. M., & Suzuki, M. (2003). Leonardo's constraint: Two opaque objects cannot be seen in the same direction. *Journal of Experimental Psychology: General*, *132*(2), 253–265.
Ono, H., Shimono, K., & Shibuta, K. (1992). Occlusion as a depth cue in the Wheatstone-Panum limiting case. *Perception & psychophysics*, *51*(1), 3–13.
Ono, H., Wade, N. J., & Lillakas, L. (2002). The pursuit of Leonardo's constraint. *Perception*, *31*(1), 83–102.
Otto, T., & Mamassian, P. (2012). Noise and Correlations in Parallel Perceptual Decision Making. *Current Biology*, 1–6.
P. L. Panum, P. L. (1858). Untersuchungen uber das Sehen mit Zwei Augen. *Kiel.*
Parker, A. J. (2007). Binocular depth perception and the cerebral cortex. *Nature Reviews Neuroscience*, *8*(5), 379–391.
Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial vision*, *10*(4), 437–442.
Pettigrew, J. D., Nikara, T., & Bishop, P. O. (1968). Binocular interaction on single units in cat striate cortex: simultaneous stimulation by single moving slit with receptive fields in correspondence. *Experimental brain research Experimentelle Hirnforschung Expérimentation cérébrale*, *6*(4), 391–410.
Pianta, M. J., & Gillam, B. J. (2003a). Paired and unpaired features can be equally effective in human depth perception. *Vision Research*, *43*(1), 1–6.
Pianta, M. J., & Gillam, B. J. (2003b). Monocular gap stereopsis: Manipulation of the outer edge disparity and the shape of the gap. *Vision Research*, *43*(18), 1937–1950.
Poggio, G. F., & Talbot, W. H. (1981). Mechanisms of static and dynamic stereopsis in foveal cortex of the rhesus monkey. *The Journal of physiology*, *315*, 469–492.
Poggio, G. F., Motter, B. C., Squatrito, S., & Trotter, Y. (1985). Responses of neurons in visual cortex (V1 and V2) of the alert macaque to dynamic random-dot stereograms. *Vision Research*, *25*(3), 397–406.
Pollard, S. B., Mayhew, J. E., & Frisby, J. P. (1985). PMF: a stereo correspondence algorithm using a disparity gradient limit. *Perception*, *14*(4), 449–470.
Portfors-Yeomans, C., & Regan, D. (1996). Cyclopean discrimination thresholds for the direction and speed of motion in depth. *Vision Research*, *36*(20), 3265–3279.
Preston, T. J., Kourtzi, Z., & Welchman, A. E. (2009). Adaptive estimation of three-dimensional structure in the human brain. *Journal of Neuroscience*, *29*(6), 1688–1698.
Preston, T. J., Li, S., Kourtzi, Z., & Welchman, A. E. (2008). Multivoxel pattern selectivity for perceptually relevant binocular disparities in the human brain. *Journal of Neuroscience*, *28*(44), 11315–11327.
Polyak, S., 1957, The Vertebrate Visual System (University of Chicago Press,

Chicago).
Rashbass, C., & Westheimer, G. (1961). Disjunctive eye movements. *Journal of Physiology*, *159*(2), 339–360.
Recanzone, G. H. (2003). Auditory influences on visual temporal rate perception. *Journal of Neurophysiology*, *89*(2), 1078–1093.
Regan, D. (1993). Binocular correlates of the direction of motion in depth. *Vision Research*, *33*(16), 2359–2360.
Regan, D., & Beverley, K. (1973a). Disparity detectors in human depth perception: Evidence for directional selectivity. *Science*, *181*(4102), 877–879.
Regan, D., & Beverley, K. I. (1973b). Electrophysiological evidence for existence of neurones sensitive to direction of depth movement. *Nature*, *246*(5434), 504–506.
Rokers, B., Cormack, L. K., & Huk, A. C. (2008). Strong percepts of motion through depth without strong percepts of position in depth. *Journal of Vision*, *8*(4), 6–6.
Rokers, B., Cormack, L. K., & Huk, A. C. (2009). Disparity- and velocity-based signals for three-dimensional motion perception in human MT+. *Nature neuroscience*, *12*(8), 1050–1055.
Rokers, B., Czuba, T. B., Cormack, L. K., & Huk, A. C. (2011). Motion processing with two eyes in three dimensions. *Journal of Vision*, *11*(2), 10–10.
Sadato, N., Pascual-Leone, A., Grafman, J., Ibañez, V., Deiber, M. P., Dold, G., & Hallett, M. (1996). Activation of the primary visual cortex by Braille reading in blind subjects. *Nature*, *380*(6574), 526–528.
Samonds, J. M., Potetz, B. R., & Lee, T. S. (2009). Cooperative and competitive interactions facilitate stereo computations in macaque primary visual cortex. *Journal of Neuroscience*, *29*(50), 15780–15795.
Schor, C. M., & Tyler, C. W. (1981). Spatio-temporal properties of Panum's fusional area. *Vision Research*, *21*(5), 683–692.
Schroeder, C. E., & Foxe, J. J. (2002). The timing and laminar profile of converging inputs to multisensory areas of the macaque neocortex. *Brain research Cognitive brain research*, *14*(1), 187–198.
Seitz, A. R., Kim, R., & Shams, L. (2006). *Sound facilitates visual learning*. *Current biology : CB* (Vol. 16, pp. 1422–1427).
Serrano-Pedraza, I. (2010). A specialization for vertical disparity discontinuities. *Journal of Vision*, *10*(3), 1–25.
Shams, L., Kamitani, Y., & Shimojo, S. (2000). Illusions. What you see is what you hear. *Nature*, *408*(6814), 788.
Shams, L., Kamitani, Y., & Shimojo, S. (2002). Visual illusion induced by sound. *Brain research Cognitive brain research*, *14*(1), 147–152.
Sheedy, J., Bailey, I., Buri, M., & Bass, E. (1986). Binocular vs. monocular task performance. *American journal of optometry and physiological optics*, *63*(10), 839–846.
Shioiri, S., Kakehi, D., Tashiro, T., & Yaguchi, H. (2009). Integration of monocular motion signals and the analysis of interocular velocity differences for the perception of motion-in-depth. *Journal of Vision*, *9*(13), 10.1–17.
Shioiri, S., Saisho, H., & Yaguchi, H. (2000). Motion in depth based on inter-ocular velocity differences. *Vision Research*, *40*(19), 2565–2572.
Spileers, W., Orban, G. A., Gulyas, B., & Maes, H. (1990). Selectivity of cat area 18 neurons for direction and speed in depth. *Journal of Neurophysiology*, *63*(4), 936–

954.
Stevenson, R. A. and James, T.W. (2009). Audiovisual integration in human superior temporal sulcus: inverse effectiveness and the neural processing of speech and object recognition, *Neuroimage 44*, 1210–1223.
Thomson, L. C. (1947). Binocular summation within the nervous pathways of the pupillary light reflex. *The Journal of physiology*, *106*(1), 59–65.
Tong, F. (2001). Competing theories of binocular rivalry: A possible resolution. *Brain and Mind*, *2*(1), 55–83.
Tong, F., Meng, M., & Blake, R. (2006). Neural bases of binocular rivalry. *Trends in cognitive sciences*, *10*(11), 502–511.
Tong, F., Nakayama, K., Vaughan, J. T., & Kanwisher, N. (1998). Binocular rivalry and visual awareness in human extrastriate cortex. *Neuron*, *21*(4), 753–759.
Treisman, A. (1962). Binocular rivalry and stereoscopic depth perception. *Quarterly Journal of Experimental Psychology*, *14*(1), 23–37.
Trivedi, H. P., & Lloyd, S. A. (1985). The role of disparity gradient in stereo vision. *Perception*, *14*(6), 685–690.
Tsao, D., Conway, B., & Livingstone, M. (2003). Receptive fields of disparity-tuned simple cells in macaque V1. *Neuron*, *38*(1), 103–114.
Tyler, C. W. (1971). Stereoscopic depth movement: two eyes less sensitive than one. *Science*, *174*(4012), 958–961.
Tyler, C. W. (1973). Stereoscopic vision: cortical limitations and a disparity scaling effect. *Science*, *181*(4096), 276.
Tyler, C. W. (1975). Spatial organization of binocular disparity sensitivity. *Vision Research*, *15*(5), 583–590.
Uka, T., Tanabe, S., Watanabe, M., & Fujita, I. (2005). Neural correlates of fine depth discrimination in monkey inferior temporal cortex. *Journal of Neuroscience*, *25*(46), 10796–10802.
Uttal, W. R., Davis, N. S., & Welke, C. (1994). Stereoscopic perception with brief exposures. *Perception & psychophysics*, *56*(5), 599–604.
Uttal, W. R., Fitzgerald, J., & Eskin, T. E. (1975). Parameters of tachistoscopic stereopsis. *Vision Research*, *15*(6), 705–712.
van Ee, R. (2009). Stochastic variations in sensory awareness are driven by noisy neuronal adaptation: evidence from serial correlations in perceptual bistability. *JOSA A*, *26*(12), 2612–2622.
Warren, P. A., Maloney, L. T., & Landy, M. S. (2002). Interpolating sampled contours in 3-D: analyses of variability and bias. *Vision Research*, *42*(21), 2431–2446.
Warren, P. A., Maloney, L. T., & Landy, M. S. (2004). Interpolating sampled contours in 3D: perturbation analyses. *Vision Research*, *44*(8), 815–832.
Watanabe, O., & Fukushima, K. (1999). Stereo algorithm that extracts a depth cue from interocularly unpaired points. *Neural networks : the official journal of the International Neural Network Society*, *12*(4-5), 569–578.
Westheimer, G. (1986). Spatial interaction in the domain of disparity signals in human stereoscopic vision. *The Journal of physiology*, *370*(1), 619–629.
Wilcox, L. M., & Lakra, D. C. (2007). Depth from binocular half-occlusions in stereoscopic images of natural scenes. *Perception*, *36*(6), 830.
Wilson, H. R. (2003). Computational evidence for a rivalry hierarchy in vision. *Proceedings of the National Academy of Sciences of the United States of America*,

100(24), 14499–14503.
Zannoli, M., Cass, J., Mamassian, P., & Alais, D. (2012). Synchronized Audio-Visual Transients Drive Efficient Visual Search for Motion-in-Depth. *PLoS ONE*, *7*(5), e37190.
Zeki, S. M. (1974). Cells responding to changing image size and disparity in the cortex of the rhesus monkey. *The Journal of physiology*, *242*(3), 827–841.